(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 February 2005 (10.02.2005)

PCT

(10) International Publication Number
WO 2005/011376 A2

(51) International Patent Classification[7]: A01N

(21) International Application Number:
PCT/US2004/024200

(22) International Filing Date: 26 July 2004 (26.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/490,087 26 July 2003 (26.07.2003) US

(71) Applicants *(for all designated States except US)*: **BIOGEN IDEC MA INC** [US/US]; 14 Cambridge Center, Cambridge, MA 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; Technology Licensing Office, Five Cambridge Center, Room NE25-230, Cambridge, MA 02142-1493 (US).

(71) Applicants and
(72) Inventors: **VAN VLIJMEN, Herman** [NL/US]; 56 Craigie Street, Somerville, MA 02143 (US). **SHERMAN, Brian, Woody, H.** [US/US]; 305 Memorial Drive, #313A, Cambridge, MA 02139 (US). **LUGOVSKOY, Alexey, Alexandrovich** [RU/US]; 24 Center Street, Woburn, MA 01808 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.** et al.; Lahive & Cockfiel, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/011376 A2

(54) Title: ALTERED ANTIBODIES HAVING IMPROVED ANTIGEN-BINDING AFFINITY

(57) Abstract: The invention relates to methods of modulating the antigen-binding affinity of an antibody by determining, using data corresponding to the structure of a complex between the antibody and an antigen in a solvent, a representation of a charge distribution of the CDRs of the antibody which minimizes electrostatic contribution to binding free energy between the antibody and the antigen in a solvent. Guided by these determinations, the antibody is accordingly modified (altered) to improve upon, *e.g.*, antibody/antigen binding by modifying at least one amino acid residue to decrease the binding free energy between the antibody and antigen when bound in a solvent.

ALTERED ANTIBODIES HAVING IMPROVED ANTIGEN-BINDING AFFINITY

Related Information

The application claims priority to U.S. provisional patent application number 60/490,087, filed on July 26, 2003, the entire contents of which are hereby incorporated by reference.

The contents of any patents, patent applications, and references cited throughout this specification are hereby incorporated by reference in their entireties.

Background of the Invention

Antibodies are exquisite, naturally occurring biological agents that play a critical role in defending the body from pathogens. Antibodies, which are also commonly referred to as immunoglobulins, contain four polypeptides: two longer polypeptides ("heavy chains") that are identical to one another and two shorter polypeptides ("light chains") that are identical to one another. The heavy chains are paired with the light chains by disulfide bonds, and the two heavy chains are similarly bound to one another to create a tetrameric structure. Moreover, the heavy and light chains each contain a variable domain and one or more constant regions: the heavy chain includes one variable domain (V_H) followed by three constant regions (C_1H, C_2H, and C_3H), and the light chain includes one variable domain (V_L) followed by a single constant region (C_L).

The variable domains of each pair of light and heavy chains form the site that comes into contact with an antigen. Both V_H and V_L have the same general structure, with four framework regions (FRs), whose sequences are relatively conserved, connected by three hypervariable or complementarity determining regions (CDRs) (*see* Kabat *et al.*, *In* "Sequences of Proteins of Immunological Interest," U.S. Department of Health and Human Services, 1983; *see also* Chothia *et al.*, *J. Mol. Biol.* 196:901-917, 1987). The four framework regions largely adopt a β-sheet conformation and the CDRs form loops connecting, and in some cases forming part of, the β-sheet structure. The CDRs of V_H and V_L are held in close proximity by the FRs, and amino acid residues within the CDRs bind the antigen. More detailed accounts of the structure of variable domains can be found in Poljak *et al.* (*Proc. Natl. Acad. Sci. USA* 70:3305-3310, 1973) Segal *et al.* (*Proc. Natl. Acad. Sci. USA* 71:4298-4302, 1974), and Marquart *et al.* (*J. Mol. Biol.*, 141:369-391, 1980).

Researchers have modified antibodies in various ways in order to study their function or to improve their utility as therapeutic agents. In some of the earliest modifications, researchers used double-stranded DNA sequences to express the V_H or V_L domains, but none of the sequence of the constant region (*see*, *e.g.*, EP-A-0 088 994; Schering Corporation). Other fragments and chimeric antibodies have also been made.

One particular type of chimera, commonly referred to as a CDR-grafted antibody, includes sequences from two antibodies that differ in species (*e.g.*, murine CDRs have been used in place of the naturally occurring CDRs in otherwise human antibodies; *see*, *e.g.*, U.S. Patent No. 5,225,539). Researchers hoped that such antibodies would be no more foreign to the human body than a genuine human antibody, but the utility of such antibodies has been restricted, at least in some cases, by a reduction in the antibody's affinity for the antigen. In an attempt to improve affinity, some of the amino acids in the FRs of CDR-grafted antibodies have been changed from those of the acceptor molecule (*e.g.*, a human antibody) to those of the antibody that donated the CDRs (*e.g.*, those of a murine antibody; *see*, *e.g.*, U.S. Patent No. 5,585,089; U.S. Patent No. 5,693,761; U.S. Patent No. 5,693,762; and U.S. Patent No. 6,180,370).

Accordingly, there remains a need for antibodies that do not provoke a strong immune response but yet bind strongly to their antigens and methods for identifying such antibodies.

Summary of the Invention

The present invention is based, in part, on the discovery that the affinity of an antibody (or an antigen-binding fragment thereof) can be improved by modifying amino acid residues within the antibody. The modifications are based, wholly or partially, on a computational analysis of electrostatic forces between the antibody and an antigen to which it binds. The computational analysis, in turn, is based on a prediction of charge distribution within the antibody that generates the electrostatic forces that influence binding between the antibody and its antigen in a solvent (*e.g.*, an aqueous solvent such as water, phosphate-buffered saline (PBS), plasma, or blood). The computational methods define the electrostatic complement (the optimal tradeoff between unfavorable desolvation energy and favorable interactions in an antigen-antibody complex) for a given target site and geometry.

In particular, the invention provides criteria or rules by which one can calculate the optimal charge distribution and associated change in binding free energy between an antibody and an antigen, when bound in a solvent, and then identify discrete residue positions for modification. Moreover, the invention provides rules which guide the selection of an appropriate modification at the identified residue position, *e.g.*, side chain chemistry, by building a subset of modifications *in silico* followed by recalculating the binding free energy and election of a preferred modification.

Thus, the invention has several advantages in that it, unlike other methods, is not restricted to mere global or pair wise alignment of charges with the presumptive conclusion that only opposite net charges between an antibody and antigen are favorable. Rather, the invention provides a more sophisticated analysis (as is appropriate given that a typical antibody comprises up to four polypeptide chains with inter and intra chain

disulfide linkages and six CDR binding surfaces as well as inter chain interfaces) for revealing the exact residue positions and side chain chemistries to be used to modify the binding-affinity of an antibody/antigen complex.

Moreover, the invention also fully accounts for the binding interactions of a antibody when bound to an antigen within a solvent.

And importantly, the invention provides for antibody modifications that alter antigen-binding which other methods would either fail to identify or dismiss as unsuitable to try.

In one aspect, the invention features a method of modulating the antigen-binding affinity of an antibody that includes the steps of providing data corresponding to the structure (*e.g.*, a three-dimensional structure) of a complex between an antibody and an antigen to which the antibody binds; determining, using the data, a representation of a charge distribution (*e.g.*, a set of multipoles or point charges) within the antibody (*e.g.*, within one or more of the CDRs) that would reduce (*i.e.*, optimize or make more negative) the electrostatic contribution to binding free energy between the antibody and the antigen; and modifying one or more amino acid residues within the antibody (*e.g.*, within one or more of the CDRs) to create a modified antibody corresponding to (or with a better correspondence to) the charge distribution (*i.e.*, the optimal charge distribution determined). The result is a charge distribution that can be used to modulate (*e.g.*, improve, alter, etc.) the interaction between an antibody and its antigen. For example, if the side chain of an amino acid residue in an optimized antibody that has a net total charge of -1, one can replace the corresponding amino acid residue in the original antibody, sometime referred to as the first antibody or parent antibody, with an amino acid residue that has a negatively charged side chain to create a modified antibody which is a variant of the parent antibody and sometimes referred to herein as a second antibody (or even a third or fourth antibody if referring to the modification of a antibody that has been previously modified and is therefore an iterative variation of the preceding antibody).

In a related aspect, the invention provides a method of modulating the antigen-binding affinity of an antibody by determining a spatial representation of an optimal charge distribution of the amino acids of the antibody and associated change in binding free energy of the antibody when bound to an antigen in a solvent; identifying at least one candidate amino acid residue position of the antibody to be modified to alter the binding free energy of the antibody when bound to the antigen; and selecting an elected amino acid residue for substitution for said amino acid position, such that upon substitution, the antigen-binding affinity of the antibody is modulated.

As described further below, once a charge distribution is determined, one or more of the amino acid residues in the antibody (*e.g.*, one or more of the residues in the CDR(s), *e.g.*, 2-10 residues or more, *e.g.*, most if not all of the CDR residues and,

optionally, only in the CDR(s)) can be modified to match, or better match, that charge distribution. For example, an amino acid residue can be replaced with another naturally occurring amino acid residue or a non-naturally occurring residue. The substitution may or may not constitute a conservative amino acid substitution. In some instances, it may be desired to alter the charge distribution by deleting or inserting one or more amino acid residues.

In some instances, for example, where the data of the structure of a complex between the antibody and the antigen is available prior to provision of the antibody, one need only know the sequence of the parent antibody (or the sequence of one or more of the CDRs of that antibody). The method can be carried out so long as one has, or can obtain, information regarding the charge distribution within an antibody-antigen complex containing a parent antibody; that information is then used to modify a modified antibody in a way that improves the modified antibody's affinity for its antigen. Alternatively, the methods of the invention can be used to alter (*e.g.*, optimize) the affinity of a fully human antibody or antigen-binding fragments containing human FRs and human CDRs, for example, affinity mature the antibody for improved antigen-binding affinity. A fully human antibody can be one obtained from human plasma (even though this is an uncommon practice) or generated *in vivo* (*e.g.*, an antibody generated in a transgenic mouse containing human immunoglobulin genes; see U.S. Patent No. 6,150,584).

In the methods of the invention, the parent and modified antibodies can be of the same or of different species (*e.g.*, the parent antibody can be a non-human antibody (*e.g.*, a murine antibody), and the modified antibody can be a human antibody). The antibodies can also be of the same, or of different, classes or subclasses. Regardless of their origin or class, portions of the sequences of the two antibodies can be identical to one another. For example, the FRs of the parent antibody can be identical to the FRs of the modified antibody. This would occur, for example, where the parent antibody is a human antibody and the modified antibody varies from the parent antibody only in that the modified antibody contains one or more non-human CDRs (*i.e.*, in the modified antibody, one or more of the original, human CDRs have been replaced with a non-human (*e.g.*, murine) CDR).

The methods of the invention can be carried out with antibodies that have the structure of a naturally occurring antibody. For example, the methods of the invention can be carried out with antibodies that have the structure of an IgG molecule (two full-length heavy chains and two full-length light chains). Thus, in some embodiments, the parent and/or modified antibody can include an Fc region of an antibody (*e.g.*, the Fc region of a human antibody). The methods of the invention can be carried out, however, with less than complete antibodies; they can be carried out with any antigen-binding fragment of an antibody including those described further below (Fab fragments, $F(ab')_2$ fragments, or single-chain antibodies (scFv)). The "fragments" can constitute minor

variations of naturally occurring antibodies. For example, an antibody fragment can include all but a few of the amino acid residues of a "complete" antibody (*e.g.*, the FR of V_H or V_L can be truncated).

Regardless of whether the method is carried out with a complete antibody or a fragment thereof, where all or part of the FR is present, the sequence of that FR can be that of a wild-type antibody. Alternatively, the FR can contain a mutation. For example, the methods of the invention can be carried out with a parent antibody that includes a framework region (*e.g.*, a human FR) that contains one or more amino acid residues that differ from the corresponding residue(s) in the wild-type FR. The mutation can be one that changes an amino acid residue to the corresponding residue in an antibody of another species. Thus, an otherwise human FR can contain a murine residue (such mutations are referred to in the art as "back mutations"). For example, framework regions of a human antibody can be "back-mutated" to the amino acid residue at the same position in a non-human antibody. Such a back-mutated antibody can be used in the present methods as the "parent" antibody, in which case the "modified" antibody can include completely human FRs. Mutations in the FRs can occur within any of FR1, FR2, FR3, and/or FR4 in either V_H or V_L (or in V_H and V_L). Up to about 10 residues or more can be mutated (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more residues in FR1, FR2, FR3, and/or FR4 can be changed from the naturally occurring residue (*e.g.*, the human residue) to another residue (*e.g.*, a donor residue, for example, murine residue, at the corresponding position)). The residues that immediately flank the CDRs are among those that can be mutated.

In one embodiment, the methods of the invention are carried out with a parent antibody that is completely non-human (*e.g.*, a murine antibody) and a modified antibody that includes a human Fc region and completely human FRs.

In certain embodiments, the relative affinities of the parent and modified antibodies (*e.g.*, the parent, modified or altered antibody of the present invention) can be such that the affinity of the modified antibody to a given antigen is at least as high as the affinity of the parent antibody to that antigen. For example, the affinity of the modified antibody to the antigen can be at least (or about) 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 5, 8, 10, 50, 10^2, 10^3, 10^4, 10^5, or 10^6, 10^7, or 10^8 times greater than the affinity of the parent antibody to the antigen (or any range or value in between).

The method may also be used lower the affinity of the antibody, for example, where it is desirable to have a lower affinity for better pharmacokinetics, antigen-binding specificity, reduced cross-talk between related antigen epitopes, and the like. For example, the affinity of the modified antibody to the antigen can be at least (or about) 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 5, 8, 10, 50, 10^2, 10^3, 10^4, 10^5, or 10^6, 10^7, or 10^8 times less than the affinity of the parent antibody to the antigen (or any range or value in between).

The methods of the invention can be iterative. An antibody generated, as described above, can be re-modeled (for example, *in silico* or empirically, *e.g.*, using experimental data) and further altered to further improve antigen binding. Thus, the steps described above can be followed by additional steps, including: obtaining data corresponding to the structure of a complex between the modified antibody and the antigen; determining, using the data (which can be referred to as "additional data" to distinguish it from the data obtained and used in the parent "round"), a representation of an additional charge distribution of the CDRs of the modified antibody which minimizes electrostatic contribution to binding free energy between the modified antibody and the antigen; and expressing a third or further modified antibody that binds to the antigen, the third antibody having a matured CDR differing from a CDR of the modified antibody by at least one amino acid, the matured CDR corresponding to the additional charge distribution. Yet additional rounds of maturation can be carried out. In the method just described, the resulting antibody would be complexed with (*i.e.* allowed to bind to) antigen and used to obtain a charge distribution that minimizes the electrostatic contribution. A fourth or further modified antibody would then be produced that would contain modifications, dictated by the charge distribution, that improve antigen binding. And so forth.

As noted above, the modified antibody (or subsequent antibodies serving in the place of the modified antibody) can contain a CDR that has been modified so that the electrostatic forces in the antibody-antigen complex are improved (or optimized). Presently, the software used to examine electrostatic forces models an optimal charge distribution and the user then determines what amino acid substitution(s) or alteration(s) would improve that distribution. Accordingly, such steps (*e.g.*, examining the modeled, optimal charge distribution and determining a sequence modification to improve antigen binding) are, or can be, part of the methods now claimed. However, as it would not be difficult to modify the software so that the program includes the selection of amino acid substitutions (or alterations), in the future, one may need only examine that output and execute the suggested change (or some variation of it, if desired).

The methods of the invention may be characterized as those that "produce" an antibody (or a fragment thereof). The term "produce" means to "make," "generate," or "design" a non-naturally occurring antibody (or fragment thereof). The antibody produced may be considered more "mature" than either of the antibodies whose sequences (*e.g.*, whose CDR(s) and FRs) were used in its construction. While the antibody produced may have a stronger affinity for an antigen, the methods of the invention are not limited to those that produce antibodies with improved affinity. For example, the methods of the invention can produce an antibody that has about the same affinity for an antigen as it did prior to being modified by the present methods. When a human antibody is modified, as described in the prior art, to contain murine CDRs, the

resulting CDR-grafted antibody can lose affinity for its antigen. Thus, for example, where the methods of the invention are applied to CDR-grafted antibodies, they are useful and successful when they prevent the loss of affinity (some or all of the loss) that would otherwise occur with a conventional CDR graft.

In addition to minimizing the electrostatic contribution to the binding free energy, the methods of the invention can further include minimizing the van der Waals or solvent accessible surface area contribution to the binding free energy. In such further computational analysis, additional amino acids in a CDR of the parent antibody may be altered to generate the modified antibody, such that the binding free energy is further reduced beyond what was achieved by solely minimizing the electrostatic contribution. As few as one and as many as 50 CDR residues may be modified in the methods and compositions of the instant invention. Most commonly, between 1 and 10 (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) amino acid residues are altered by the methods and compositions of the instant invention.

Antibodies produced by any of the methods of the invention are also within the scope of the invention, pharmaceutical compositions containing those antibodies, as well as nucleic acids encoding such antibodies. The present invention also includes vectors that express the modified antibodies (or polypeptides or fragments thereof) found by the methods described above. These vectors can be used to transform cell lines, and such transformed (*e.g.* transfected) cells are within the scope of the invention.

The details of one or more embodiments of the invention are set forth in the description below. Other features, objects, and advantages of the invention will be apparent from the description and the claims.

Brief Description of the Figures

Figure 1 illustrates geometries for modeling the binding interactions between an antibody, or antigen-binding fragment thereof, and an antigen, when bound in a solvent (top panel). In particular, the boundary-value problem which comprises a determination of the charge distribution in a spherical region of radius R with a dielectric constant $\in_1$, surrounded by solvent with a dielectric constant $\in_2$ as well as other geometries of the antibody-antigen interface (bottom panel, see also text, *infra*).

Figure 2 depicts nucleotide (SEQ ID NOs: 1, 3) and polypeptide (SEQ ID NOs: 2, 4) sequences for 5c8 heavy variable and light variable chain domains.

Detailed Description of the Invention

In order to provide a clear understanding of the specification and claims, the following definitions are conveniently provided below.

Definitions

The term "structure", or "structural data", as used herein, includes the known, predicted and/or modeled position(s) in three-dimensional space that are occupied by the atoms, molecules, compounds, amino acid residues and portions thereof, and macromolecules and portions thereof, of the invention, and, in particular, an antibody bound to an antigen in a solvent. A number of methods for identifying and/or predicting structure at the molecular/atomic level can be used such as X-ray crystallography, NMR structural modeling, and the like.

The term "binding affinity", as used herein, includes the strength of a binding interaction and therefore includes both the actual binding affinity as well as the apparent binding affinity. The actual binding affinity is a ratio of the association rate over the disassociation rate. Therefore, conferring or optimizing binding affinity includes altering either or both of these components to achieve the desired level of binding affinity. The apparent affinity can include, for example, the avidity of the interaction. For example, a bivalent altered variable region binding fragment can exhibit altered or optimized binding affinity due to its valency. Binding affinities may also be modeled, with such modeling contributing to selection of residue alterations in the methods of the current invention.

The term "binding free energy" or "free energy of binding", as used herein, includes its art-recognized meaning, and, in particular, as applied to antibody-antigen interactions in a solvent. Reductions in binding free energy enhance antibody-antigen affinities, whereas increases in binding free energy reduce antibody-antigen affinities.

The phrase "spatial representation of an optimal charge distribution", as used herein, includes modeling the charge distribution for an antibody or antibody-antigen

complex, wherein the electrostatic contribution to free energy of the antibody when bound to antigen is optimized (minimized), as compared to the known and/or modeled representation of charge distribution of the parent antibody and/or parent antibody when bound to antigen. The modeling of optimal charge distribution can be arrived at by an *in silico* process that incorporates the known and/or modeled structure(s) of an antibody and/or antibody-antigen complex as an input. Response continuum modeling (*e.g.*, the linearized Poisson-Boltzmann equation) can be employed to express the electrostatic binding free energy of the antigen-antibody complex in a solvent as a sum of antibody desolvation, antibody-antigen interaction, and antigen desolvation terms. This *in silico* process is characterized by the ability to incorporate monopole, dipolar, and quadrupolar terms in representing charge distributions within the modeled charge distributions of the invention, and allows for extensive assessment of solvation/desolvation energies for antibody residues during transition of the antibody between unbound and bound states. The process of modeling the spatial representation of an optimal charge distribution for an antibody-antigen complex may additionally incorporate modeling of van der Waals forces, solvent accessible surface area forces, *etc.*

The term "solvent", as used herein, includes its broadest art-recognized meaning, referring to any liquid in which an antibody of the instant invention is dissolved and/or resides.

The term "antibody", as used herein, includes monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, multispecific antibodies (*e.g.*, bispecific antibodies), chimeric antibodies, CDR-grafted antibodies, humanized antibodies, human antibodies and antigen-binding fragments thereof, for example, an antibody light chain (VL), an antibody heavy chain (VH), a single chain antibody (scFv), a F(ab')2 fragment, a Fab fragment, an Fd fragment, an Fv fragment, and a single domain antibody fragment (DAb).

The term "antigen", as used herein, includes an entity (*e.g.*, a proteinaceous entity or peptide) to which an antibody specifically binds, and includes, *e.g.*, a predetermined antigen to which both a parent antibody and modified antibody as herein defined bind. The target antigen may be polypeptide, carbohydrate, nucleic acid, lipid, hapten, or other naturally occurring or synthetic compound. Preferably, the target antigen is a polypeptide.

The term "CDR", as used herein, includes the complementarity determining regions as described by, for example Kabat , Chothia, or MacCallum *et al.*, (see, *e.g.*, Kabat *et al.*, *In* "Sequences of Proteins of Immunological Interest," U.S. Department of Health and Human Services, 1983; Chothia *et al.*, *J. Mol. Biol.* 196:901-917, 1987; and MacCallum *et al.*, *J. Mol. Biol.* 262:732-745 (1996); the contents of which are incorporated herein in their entirety).

The amino acid residue positions which typically encompass the CDRs as

described by each of the above cited references are set forth below for comparison.

Table of CDR Definitions

	Kabat	Chothia	MacCallum
V_H CDR1	31-35	26-32	30-35
V_H CDR2	50-65	53-55	47-58
V_H CDR3	95-102	96-101	93-101
V_L CDR1	24-34	26-32	30-36
V_L CDR2	50-56	50-52	46-55
V_L CDR3	89-97	91-96	89-96

The term "variable region", as used herein, includes the amino terminal portion of an antibody which confers antigen binding onto the molecule and which is not the constant region. The term is intended to include functional fragments, for example, antigen-binding fragments, which maintain some or all of the binding function of the whole variable region.

The term "framework region", as used herein, includes the antibody sequence that is between and separates the CDRs. Therefore, a variable region framework is between about 100-120 amino acids in length but is intended to reference only those amino acids outside of the CDRs. For the specific example of a heavy chain variable region and for the CDRs as defined by Kabat *et al.*, framework region 1 corresponds to the domain of the variable region encompassing amino acids 1-30; region 2 corresponds to the domain of the variable region encompassing amino acids 36-49; region 3 corresponds to the domain of the variable region encompassing amino acids 66-94, and region 4 corresponds to the domain of the variable region from amino acids 103 to the end of the variable region. The framework regions for the light chain are similarly separated by each of the light claim variable region CDRs. Similarly, using the definition of CDRs by Chothia *et al.* or McCallum *et al.* the framework region boundaries are separated by the respective CDR termini as described above.

The term terms "modified" or "altered", as used herein, include antibodies or antigen-binding fragments thereof, that contain one or more amino acid changes in, for example, a CDR(s), a framework region(s), or both as compared to the parent amino acid sequence at the changed position. A modified or altered antibody typically has one or more residues which has been substituted with another amino acid residue, related side chain chemistry thereof, or one or more amino acid residue insertions or deletions.

The term "parent antibody", "original antibody", "starting antibody", "wild-type", or "first antibody", as used herein, includes any antibody for which modification of antibody-antigen binding affinity by the methods of the instant invention is desired. Thus, the parent antibody represents the input antibody on which the methods of the instant invention are performed. The parent polypeptide may comprise a native sequence (*i.e.* a naturally occurring) antibody (including a naturally occurring allelic variant), or an

antibody with pre-existing amino acid sequence modifications (such as insertions, deletions and/or other alterations) of a naturally occurring sequence. The parent antibody may be a monoclonal, chimeric, CDR-grafted, humanized, or human antibody.

The terms "antibody variant", "modified antibody", "antibody containing a modified amino acid", "mutant", or "second antibody", "third antibody", *etc.*, as used herein, include an antibody which has an amino acid sequence which differs from the amino acid sequence of a parent antibody. Preferably, the antibody variant comprises a heavy chain variable domain or a light chain variable domain having an amino acid sequence which is not found in nature. Such variants necessarily have less than 100% sequence identity or similarity with the parent antibody. In a preferred embodiment, the antibody variant will have an amino acid sequence from about 75% to less than 100% amino acid sequence identity or similarity with the amino acid sequence of either the heavy or light chain variable domain of the parent antibody, more preferably from about 80% to less than 100%, more preferably from about 85% to less than 100%, more preferably from about 90% to less than 100%, and most preferably from about 95% to less than 100%. Identity or similarity with respect to this sequence is defined herein as the percentage of amino acid residues in the candidate sequence that are identical (*i.e.* same residue) with the parent antibody residues, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Typically, N-terminal, C-terminal, or internal extensions, deletions, or insertions into the antibody sequence outside of the variable domain are not construed as affecting sequence identity or similarity. The antibody variant is generally one which comprises one or more amino acid alterations in or adjacent to one or more hypervariable regions thereof. The modified antibodies of the present invention may either be expressed, or alternatively, may be modeled *in silico*.

The phrase "candidate amino acid residue position", as used herein, includes an amino acid position identified within an antibody of the present invention, wherein the substitution of the candidate amino acid is modeled, predicted, or known to impact charge distribution of the antibody upon alteration, deletion, insertion, or substitution with another amino acid.

The term "elected amino acid", as used herein, refers to an amino acid residue(s) that has been selected by the methods of the present invention for substitution as a replacement amino acid at the candidate amino acid position within the antibody. Substitution of the candidate amino acid residue position with the elected amino acid residue may either reduce or increase the electrostatic contribution to binding free energy of the antibody-antigen complex.

The terms "amino acid alteration" or "alteration for said amino acid", as used herein, include refers to a change in the amino acid sequence of a predetermined amino acid sequence. Exemplary alterations include insertions, substitutions, and deletions.

The term "amino acid modification", as used herein, includes the replacement of an existing amino acid residue side chain chemistry in a predetermined amino acid sequence with another different amino acid residue side chain chemistry, by, for example, amino acid substitution. Individual amino acid modifications of the instant invention are selected from any one of the following: (1) the set of amino acids with nonpolar sidechains, *e.g.*, Ala, Cys, Ile, Leu, Met, Phe, Pro, Val, (2) the set of amino acids with negatively charged side chains, *e.g.*, Asp, Glu, (3) the set of amino acids with positively charged sidechains, *e.g.*, Arg, His, Lys, and (4) the set of amino acids with uncharged polar sidechains, *e.g.*, Asn, Cys, Gln, Gly, His, Met, Phe, Ser, Thr, Trp, Tyr, to which are added Cys, Gly, Met and Phe.

The term "naturally occurring amino acid residue", as used herein, includes one encoded by the genetic code, generally selected from the group consisting of: alanine (Ala); arginine (Arg); asparagine (Asn); aspartic acid (Asp); cysteine (Cys); glutamine (Gln); glutamic acid (Glu); glycine (Gly); histidine (His); isoleucine (Ile): leucine (Leu); lysine (Lys); methionine (Met); phenylalanine (Phe); proline (Pro); serine (Ser); threonine (Thr); tryptophan (Trp); tyrosine (Tyr); and valine (Val).

The term "non-naturally occurring amino acid residue". as used herein, includes an amino acid residue other than those naturally occurring amino acid residues listed above, which is able to covalently bind adjacent amino acid residues(s) in a polypeptide chain. Examples of non-naturally occurring amino acid residues include norleucine, omithine, norvaline, homoserine and other amino acid residue analogues such as those described in Ellman *et al. Meth. Enzym.* 202:301-336 (1991). To generate such non-naturally occurring amino acid residues, the procedures of Noren *et al. Science* 244:182 (1989) and Ellman *et al.*, *supra*, can be used. Briefly, these procedures involve chemically activating a suppressor tRNA with a non-naturally occurring amino acid residue followed by in vitro transcription and translation of the RNA.

The term "exposed" amino acid residue, as used herein, includes one in which at least part of its surface is exposed, to some extent, to solvent when present in a polypeptide (*e.g.*, an antibody or polypeptide antigen) in solution. Preferably, the exposed amino acid residue is one in which at least about one third of its side chain surface area is exposed to solvent. Various methods are available for determining whether a residue is exposed or not, including an analysis of a molecular model or structure of the polypeptide.

The term "treatment" refers to both therapeutic treatment and prophylactic or preventative measures. Those in need of treatment include those already with the disorder as well as those in which the disorder is to be prevented.

The term "disorder or disease" is any condition that would benefit from treatment with the antibody variant. This includes chronic and acute disorders or diseases including those pathological conditions which predispose the mammal to the disorder in question.

The terms "cell", "cell line", "cell culture", or "host cell", as used herein, includes "transformants", "transformed cells", or "transfected cells" and progeny thereof. Host cells within the scope of the invention include prokaryotic cells such as *E. coli*, lower eukaryotic cells such as yeast cells, insect cells, and higher eukaryotic cells such as vertebrate cells, for example, mammalian cells, *e.g.*, Chinese hamster ovary cells and NS0 myeloma cells.

Detailed Description

Overview

The methods described herein can be used to obtain an optimized antibody (or an antigen-binding fragment thereof). Based on a computational analysis, positions are identified within any given antibody where there is a difference (the larger the difference, the more significant it can be) between the charge distribution in an optimized antibody-antigen complex and that in an original antibody-antigen complex. Such differences in charge distribution are also associated with changes in binding free energy of the antibody when bound to the antigen in a solvent. The amino acid residue at such a position can then be changed so that the electrostatic forces in the original antibody more nearly approach (or in alternative embodiments, are more divergent from) those in the optimized antibody, thereby modulating binding free energy of the antibody when bound to an antigen in a solvent. Changes to the antibody are introduced according to a set of discrete criteria or rules as described herein.

Rules for Modifying Antibodies for Improved Function

The rules of the invention can be applied as follows. To modulate the antigen-binding affinity of an antibody, for example, to improve or restore such binding, basic sequence and/or structural data is first acquired. Electrostatic charge optimization techniques are then applied to suggest improved-affinity mutants. Typically, an electrostatic charge optimization is first used to determine the position(s) of the CDR residue(s) that are sub-optimal for binding (Lee and Tidor, *J. Chem. Phys.* 106:8681-8690, 1997; Kangas and Tidor, *J. Chem. Phys.* 109:7522-7545, 1998). Then, one or more CDR mutations (*i.e.*, modifications) is subjected to further computational analysis. Based on these calculations, the binding affinity is then determined for a subset of modified antibodies having one or more modifications according to the rules of the invention.

Using a continuum electrostatics model, an electrostatic charge optimization can be performed on each side chain of the amino acids in the CDRs of the antibody. A charge optimization gives charges at atom centers but does not always yield actual mutation(s). Accordingly, a round of charge optimizations can be performed with various constraints imposed to represent natural side chain characteristics at the positions of interest. For example, an optimization can be performed for a net side chain charge of –1,

0, and +1 with the additional constraint that no atom's charge exceeded a particular value, *e.g.*, 0.85 electron charge units. Candidate amino acid side chain positions, and residue modifications at these positions, are then determined based on the potential gain in electrostatic binding free energy observed in the optimizations.

Binding free energy difference (in kcal/mol) in going from the native residue to a completely uncharged sidechain isostere, *i.e.*, a residue with the same shape but no charges or partial charges on the atoms can be calculated. Negative numbers indicate a predicted increase of binding affinity. Optimal charge distribution wherein the net side chain charge is +1, 0, or -1 can be used to calculate the binding free energy difference.

In those instances in which binding free energy difference is favorable ($\Delta G < -0.25$ kcal/mol) and associated with a transition from the native residue to a completely uncharged side chain isostere, *i.e.*, a residue with the same shape but no charges or partial charges on the atoms, modifications from the set of amino acids with nonpolar sidechains, *e.g.*, Ala, Cys, Ile, Leu, Met, Phe, Pro, Val are selected.

Where the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of –1 is favorable ($\Delta G < -0.25$ kcal/mol), modifications from the set of amino acids with negatively charged side chains, *e.g.*, Asp, Glu are selected.

Similarly, where the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of +1 is favorable ($\Delta G < -0.25$ kcal/mol), modifications from the set of amino acids with positively charged sidechains, *e.g.*, Arg, His, Lys are selected.

Finally, in those cases where the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of 0 is favorable ($\Delta G < -0.25$ kcal/mol), modifications from the set of amino acids with uncharged polar sidechains, *e.g.*, Asn, Cys, Gln, Gly, His, Met, Phe, Ser, Thr, Trp, Tyr, to which are added Cys, Gly, Met and Phe are selected.

As described herein, the designed modified antibodies can be built *in silico* and the binding energy recalculated. Modified side chains can be built by performing a rotamer dihedral scan in CHARMM, using dihedral angle increments of 60 degrees, to determine the most desirable position for each side chain. Binding energies are then calculated for the wild type (parent) and mutant (modified) complexes using the Poisson-Boltzmann electrostatic energy and additional terms for the van der Waals energy and buried surface area.

Results from these computational modification calculations are then reevaluated as needed, for example, after subsequent reiterations of the method either *in silico* or informed by additional experimental structural/functional data.

The rules allow for several predictions to be made which can be categorized as follows:

1) modifications at the interaction interface involving residues on the antibody that become partially buried upon binding (interactions are improved by making hydrogen bonds with the antigen);

2) modifications of polar residues on the antibody that become buried upon binding and thus pay a desolvation penalty but do not make any direct electrostatic interactions with the antigen (improvements are usually made by modifying to a hydrophobic residue with similar shape to the wild-type residue or by adding a residue that can make favorable electrostatic interactions); and

3) modifications of surface residues on the antibody that are in regions of uncomplementary potentials. These modifications are believed to improve long-range electrostatic interactions between the antibody and antigen without perturbing packing interactions at the binding interface.

Thus practiced, the rules of the invention allow for the successful prediction of affinity altering, *e.g.*, enhancing, side chain modifications. These findings can be classified into three general classes of modifications. The first type of modification involves residues at the interface across from a charged group on the antigen capable of making a hydrogen bond; the second type involves buried polar residues that pay a desolvation penalty upon binding but do not make back electrostatic interactions; and the third type involves long-range electrostatic interactions.

The first type of modification is determined by inspection of basic physical/chemical considerations, as these residues essentially make hydrogen bonds with unsatisfied hydrogen partners of the antigen. Unlike other methods, the rules of the invention allowed for surprising residue modifications in which the cost of desolvation is allowed to outweigh the beneficial interaction energy.

The second type of modification represents still another set of modifications, as the energy gained is primarily a result of eliminating an unfavorable desolvation while maintaining non-polar interactions.

The third type of modification concerns long-range interactions that show potential for significant gain in affinity. These types of modifications are particularly interesting because they do not make direct contacts with the antigen and, therefore, pose less of a perturbation in the delicate interactions at the antibody-antigen interface.

Accordingly, when the desired side chain chemistries are determined for the candidate amino acid position(s) according to the rules, the residue position(s) is then modified or altered, *e.g.*, by substitution, insertion, or deletion, as further described herein.

In addition to the above rules for antibody modification, it is noted that certain determinations, *e.g.*, solvent effects can be factored into initial (and subsequent) calculations of optimal charge distributions.

Obtaining an Antibody or Antigen-Binding Fragment Thereof

The methods of the invention that are aimed at generating a non-naturally occurring antibody (or an antigen-binding fragment thereof) can, but do not necessarily, begin by obtaining an antibody. That antibody may be referred to herein as a "parent" antibody or sometimes as a "first" antibody, and it can be used to obtain information that will allow one to modify or alter one or more amino acid residues either within that antibody (*i.e.*, within the parent antibody) or within a modified or altered antibody having a sequence that is similar to, or that contains portions of, the sequence of the parent antibody. As described herein, for example, one or more of the CDRs (or portions thereof) of a parent antibody, can be replaced with the corresponding CDR(s) of the modified antibody by standard genetic engineering techniques to accomplish the so-called CDR graft or transplant. Accordingly, the method can begin with a mammalian monoclonal or polyclonal antibody (*e.g.*, murine or primate), chimeric, CDR-grafted, humanized, or human antibody.

The parent antibodies can be obtained from art-recognized sources or produced according to art-recognized technologies. For example, the parent antibody can be a CDR-grafted or humanized antibody having CDR regions derived from another source or species, *e.g.*, murine.

The parent antibody or any of the modified antibodies of the invention can be in the format of a monoclonal antibody. Methods for producing monoclonal antibodies are known in the art (*see, e.g.*, Kohler and Milstein, *Nature* 256:495-497, 1975), as well as techniques for stably introducing immunoglobulin-encoding DNA into myeloma cells (*see, e.g.*, Oi *et al.*, *Proc. Natl. Acad. Sci. USA* 80:825-829, 1983; Neuberger, *EMBO J.* 2:1373-1378, 1983; and Ochi *et al.*, *Proc. Natl. Acad. Sci. USA* 80:6351-6355, 1983). These techniques, which include *in vitro* mutagenesis and DNA transfection, allow for the construction of recombinant immunoglobulins; these techniques can be used to produce the parent and modified antibodies used in the methods of the invention or to produce the modified antibodies that result from those methods. Alternatively, the parent antibodies can be obtained from a commercial supplier. Antibody fragments (scFvs and Fabs) can also be produced in *E. coli* (production methods and cellular hosts are described further below).

The parent antibody or any of the modified antibodies of the invention can be an antibody of the IgA, IgD, IgE, IgG, or IgM class.

As noted above, the methods of the invention can be applied to more than just tetrameric antibodies (*e.g.*, antibodies having the structure of an immunoglobulin of the G class (an IgG)). For example, the methods of modifying an antibody can be carried out with antigen-binding fragments of any antibody as well. The fragments can be recombinantly produced and engineered, synthesized, or produced by digesting an antibody with a proteolytic enzyme. For example, the fragment can be an Fab fragment;

digestion with papain breaks the antibody at the region, before the inter-chain (*i.e.*, V_H-V_H) disulphide bond, that joins the two heavy chains. This results in the formation of two identical fragments that contain the light chain and the V_H and C_H1 domains of the heavy chain. Alternatively, the fragment can be an $F(ab')_2$ fragment. These fragments can be created by digesting an antibody with pepsin, which cleaves the heavy chain after the inter-chain disulphide bond, and results in a fragment that contains both antigen-binding sites. Yet another alternative is to use a "single chain" antibody. Single-chain Fv (scFv) fragments can be constructed in a variety of ways. For example, the C-terminus of V_H can be linked to the N-terminus of V_L. Typically, a linker (*e.g.*, $(GGGGS)_4$) is placed between V_H and V_L. However, the order in which the chains can be linked can be reversed, and tags that facilitate detection or purification (*e.g.*, Myc-, His-, or FLAG-tags) can be included (tags such as these can be appended to any antibody or antibody fragment of the invention; their use is not restricted to scFv). Accordingly, and as noted below, tagged antibodies are within the scope of the present invention. In alternative embodiments, the antibodies used in the methods described herein, or generated by those methods, can be heavy chain dimers or light chain dimers. Still further, an antibody light or heavy chain, or portions thereof, for example, a single domain antibody (DAb), can be used.

As the methods of the invention can be iterative, the parent antibody may not be a naturally occurring antibody. As the process of modifying an antibody can be repeated as many times as necessary, the starting antibody (or antigen-binding fragment thereof) can be wholly non-human or an antibody containing human FRs and non-human (*e.g.*, murine) CDRs. That is, the "parent" antibody can be a CDR-grafted antibody that is subjected to the methods of the invention in order to improve the affinity of the antibody, *i.e.*, affinity mature the antibody. As noted above, the affinity may only be improved to the extent that it is about the same as (or not significantly worse than) the affinity of the naturally occurring human antibody (the FR-donor) for its antigen. Thus, the "parent" antibody may, instead, be an antibody created by one or more earlier rounds of modification, including an antibody that contains sequences of more than one species (*e.g.*, human FRs and non-human CDRs). The methods of the invention encompass the use of a "parent" antibody that includes one or more CDRs from a non-human (*e.g.*, murine) antibody and the FRs of a human antibody. Alternatively, the parent antibody can be completely human.

Where the structure is available, of course, one may begin the computational analysis with that structure (rather than creating it again).

The Method of the Invention Informed by Antibody-Antigen Structural Data

Proteins are known to fold into three-dimensional structures that are dictated by the sequences of their amino acids and by the solvent in which a given protein (or protein-

containing complex) is provided. The three-dimensional structure of a protein influences its biological activity and stability, and that structure can be determined or predicted in a number of ways. Generally, empirical methods use physical biochemical analysis. Alternatively, tertiary structure can be predicted using model building of three-dimensional structures of one or more homologous proteins (or protein complexes) that have a known three-dimensional structure. X-ray crystallography is perhaps the best-known way of determining protein structure (accordingly, the term "crystal structure" may be used in place of the term "structure"), but estimates can also be made using circular dichroism, light scattering, or by measuring the absorption and emission of radiant energy. Other useful techniques include neutron diffraction and nuclear magnetic resonance (NMR). All of these methods are known to those of ordinary skill in the art, and they have been well described in standard textbooks (*see, e.g., Physical Chemistry*, 4th Ed., W.J. Moore, Prentiss-Hall, N.J., 1972, or *Physical Biochemistry*, K.E. Van Holde, Prentiss-Hall, N.J., 1971)) and numerous publications. Any of these techniques can be carried out to determine the structure of an antibody, or antibody -antigen-containing complex, which can then be analyzed according to the methods of the present invention and, *e.g.*, used to inform one or more steps of the method of the invention.

Similarly, these and like methods can be used to obtain the structure of an antigen bound to an antibody fragment, including a fragment consisting of, *e.g.*, a single-chain antibody, Fab fragment, *etc.* Methods for forming crystals of an antibody, an antibody fragment, or scFv-antigen complex have been reported by, for example, van den Elsen *et al.* (*Proc. Natl. Acad. Sci. USA* 96:13679-13684, 1999, which is expressly incorporated by reference herein).

Computational Analysis

The basic computational formulae used in carrying out the methods of the invention are provided in, *e.g.*, U.S. Patent No. 6,230,102, the contents of which are hereby incorporated by reference in the present application in their entirety.

As noted above, antibodies are altered (or "modified") according to the results of a computational analysis of electrostatic forces between the antibody and an antigen to which it binds, preferably, in accordance to the discrete criteria or rules of the invention described herein. The computational analysis allows one to predict the optimal charge distribution within the antibody, and one way to represent the charge distribution in a computer system is as a set of multipoles. Alternatively, the charge distribution can be represented by a set of point charges located at the positions of the atoms of the antibody. Once a charge distribution is determined (preferably, an optimal charge distribution), one can modify the antibody to match, or better match, that charge distribution.

The computational analysis can be mediated by a computer-implemented process that carries out the calculations described in U.S. Patent No. 6,230,102. The computer

program is adapted herein to consider the real world context of antigen-antibody binding (and unlike other methods, this methods of the invention take into account, *e.g.*, solvent, long-range electrostatics, and dielectric effects in the binding between an antibody and its antigen in a solvent). The process is used to identify modifications to the antibody structure that will achieve a charge distribution on the "matured" antibody that minimizes the electrostatic contribution to binding free energy between the matured antibody and its antigen (compared to that of the unmodified ("starting" or "parent") antibody. As is typical, the computer system (or device(s)) that performs the operations described here (and in more detail in U.S. Patent No. 6,230,102) will include an output device that displays information to a user (*e.g.*, a CRT display, an LCD, a printer, a communication device such as a modem, audio output, and the like). In addition, instructions for carrying out the method, in part or in whole, can be conferred to a medium suitable for use in an electronic device for carrying out the instructions. Thus, the methods of the invention are amendable to a high throughput approach comprising software (*e.g.*, computer-readable instructions) and hardware (*e.g.*, computers, robotics, and chips). The computer-implemented process is not limited to a particular computer platform, particular processor, or particular high-level programming language.

A useful process is set forth in Appendix A (U.S. Patent No. 6,230,102) and a more detailed exposition is provided in Appendix B (Lee and Tidor (*J. Chem. Phys.* 106:8681-8690, 1997; each of which is expressly incorporated herein by reference).

Analysis of Affinity

Affinity, avidity, and/or specificity can be measured in a variety of ways. Generally, and regardless of the precise manner in which affinity is defined or measured, the methods of the invention improve antibody affinity when they generate an antibody that is superior in any aspect of its clinical application to the antibody (or antibodies) from which it was made (for example, the methods of the invention are considered effective or successful when a modified antibody can be administered at a lower dose or less frequently or by a more convenient route of administration than an antibody (or antibodies) from which it was made).

More specifically, the affinity between an antibody and an antigen to which it binds can be measured by various assays, including, *e.g.*, a BiaCore assay or the KinExA™ 3000 assay (available from Sapidyne Instruments (Boise, ID)). The latter assay was used to measure the affinity of AQC2 scFv mutants for the VLA1 I domain (see the Examples below). Briefly, sepharose beads are coated with antigen (in the Examples below, the antigen is a VLA1 I-domain protein, but the antigen used in the methods of the invention can be any antigen of interest (*e.g.*, a cancer antigen; a cell surface protein or secreted protein; an antigen of a pathogen (*e.g.*, a bacterial or viral antigen (*e.g.*, an HIV antigen, an influenza antigen, or a hepatitis antigen)), or an allergen)

by covalent attachment. (It is understood, however, that the methods described here are generally applicable; they are not limited to the production of antibodies that bind any particular antigen or class of antigens.)

Those of ordinary skill in the art will recognize that determining affinity is not always as simple as looking at a single, bottom-line figure. Since antibodies have two arms, their apparent affinity is usually much higher than the intrinsic affinity between the variable region and the antigen (this is believed to be due to avidity). Intrinsic affinity can be measured using scFv or Fab fragments.

Chimeric Antibodies and Antibody Fragments

The term "chimeric antibody" is used to describe a protein comprising at least an antigen-binding portion of an immunoglobulin molecule that is attached by, for example, a peptide bond or peptide linker, to a heterologous protein or a peptide thereof. The "heterologous" protein can be a non-immunoglobulin or a portion of an immunoglobulin of a different species, class or subclass.

There are numerous processes by which such antibodies can be made. For example, one can prepare an expression vector including a promoter that is operably linked to a DNA sequence that encodes at least V_H or V_L and a sequence that encodes the heterologous protein (or a peptide thereof (the peptide being of a sufficient length that it can be recognized as a non-immunoglobulin molecule (*i.e.*, a peptide having no substantial sequence identity to an immunoglobulin))). If necessary, or desired, one can prepare a second expression vector including a promoter that is operably linked to a DNA sequence that encodes the complementary variable domain (*i.e.*, where the parent expression vector encodes V_H, the second expression vector encodes V_L and *vice versa*). A cell line (*e.g.*, an immortalized mammalian cell line) can then be transformed with one or both of the expression vectors and cultured under conditions that permit expression of the chimeric variable domain or chimeric antibody (*see*, *e.g.*, International Patent Application No. PCT/GB85/00392 to Neuberger *et. al.*). While Neuberger *et al.* produced chimeric antibodies in which complete variable domains were encoded by the parent expression vector, this method can be used to express the modified antibodies of the present invention, antibodies containing full-length heavy and light chains, or fragments thereof (*e.g.*, the Fab, $F(ab')_2$, or scFv fragments described herein). The methods are not limited to expression of chimeric antibodies.

The antibodies produced by the methods described herein can be labeled just as any other antibody can be labeled. Accordingly, the invention encompasses antibodies produced by the present methods that are labeled with detectable labels such as a radioactive label (*e.g.*, P^{32} or S^{35}), an enzyme (*e.g.*, horseradish peroxidase, chloramphenicol acetyltransferase (CAT), β-galactosidase (β-gal), or the like), a chromophore or a fluorophore including a quantum dot. The labeled antibodies can be

used to carry out diagnostic procedures (many diagnostic assays rely on detection of a protein antigen (such as PSA)) in a variety of cell or tissue types. For imaging procedures, *in vitro* or *in vivo*, the altered antibodies produced by the methods described herein can be labeled with additional agents, such as NMR contrasting agents, X-ray contrasting agents, or quantum dots. Methods for attaching a detectable agent to polypeptides, including antibodies or fragments thereof, are known in the art. The antibodies can also be attached to an insoluble support (such as a bead, a glass or plastic slide, or the like).

Construction of Modified Antibodies

Once the sequence of an antibody (*e.g.*, a CDR-grafted or otherwise modified or "humanized" antibody) has been decided upon, that antibody can be made by techniques well known in the art of molecular biology. More specifically, recombinant DNA techniques can be used to produce a wide range of polypeptides by transforming a host cell with a nucleic acid sequence (*e.g.*, a DNA sequence that encodes the desired protein products (*e.g.*, a modified heavy or light chain; the variable domains thereof, or other antigen-binding fragments thereof)).

More specifically, the methods of production can be carried out as described above for chimeric antibodies. The DNA sequence encoding, for example, an altered variable domain can be prepared by oligonucleotide synthesis. The variable domain can be one that includes the FRs of a human acceptor molecule and the CDRs of a donor, *e.g.*, murine, either before or after one or more of the residues (*e.g.*, a residue within a CDR) has been modified to facilitate antigen binding. This is facilitated by determining the framework region sequence of the acceptor antibody and at least the CDR sequences of the donor antibody. Alternatively, the DNA sequence encoding the altered variable domain may be prepared by primer directed oligonucleotide site-directed mutagenesis. This technique involves hybridizing an oligonucleotide coding for a desired mutation with a single strand of DNA containing the mutation point and using the single strand as a template for extension of the oligonucleotide to produce a strand containing the mutation. This technique, in various forms, is described by, *e.g.*, Zoller and Smith (*Nuc. Acids Res.* 10:6487-6500, 1982), Norris *et al.* (*Nuc. Acids Res.* 11:5103-5112, 1983), Zoller and Smith (*DNA* 3:479-488, 1984), and Kramer *et al.* (*Nuc. Acids Res.* 10:6475-6485, 1982).

Other methods of introducing mutations into a sequence are known as well and can be used to generate the altered antibodies described herein (*see*, *e.g.*, Carter *et al.*, *Nuc. Acids Res.* 13:4431-4443, 1985). The oligonucleotides used for site-directed mutagenesis can be prepared by oligonucleotide synthesis or isolated from DNA coding for the variable domain of the donor antibody by use of suitable restriction enzymes.

Host Cells and Cell Lines for Expression of the Modified Antibodies

Either the parent antibodies or modified antibodies as described herein (whether in a final form or an intermediate form) can be expressed by host cells or cell lines in culture. They can also be expressed in cells *in vivo*. The cell line that is transformed (*e.g.*, transfected) to produce the altered antibody can be an immortalised mammalian cell line, such as those of lymphoid origin (*e.g.*, a myeloma, hybridoma, trioma or quadroma cell line). The cell line can also include normal lymphoid cells, such as B-cells, that have been immortalized by transformation with a virus (*e.g.*, the Epstein-Barr virus).

Although typically the cell line used to produce the altered antibody is a mammalian cell line, cell lines from other sources (such as bacteria and yeast) can also be used. In particular, *E. coli*-derived bacterial strains can be used, especially, *e.g.*, phage display.

Some immortalized lymphoid cell lines, such as myeloma cell lines, in their normal state, secrete isolated Ig light or heavy chains. If such a cell line is transformed with a vector that expresses an altered antibody, prepared during the process of the invention, it will not be necessary to carry out the remaining steps of the process, provided that the normally secreted chain is complementary to the variable domain of the Ig chain encoded by the vector prepared earlier.

If the immortalised cell line does not secrete or does not secrete a complementary chain, it will be necessary to introduce into the cells a vector that encodes the appropriate complementary chain or fragment thereof.

In the case where the immortalised cell line secretes a complementary light or heavy chain, the transformed cell line may be produced for example by transforming a suitable bacterial cell with the vector and then fusing the bacterial cell with the immortalised cell line (*e.g.*, by spheroplast fusion). Alternatively, the DNA may be directly introduced into the immortalised cell line by electroporation.

Pharmaceutical Formulations and Their Uses

In prophylactic applications, pharmaceutical compositions or medicaments are administered to a subject suffering from a disorder in an amount sufficient to eliminate or reduce the risk, lessen the severity, or delay the outset of the disorder, including biochemical, histologic and/or behavioral symptoms of the disorder, its complications and intermediate pathological phenotypes presenting during development of the disorder. In therapeutic applications, compositions or medicaments are administered to a subject suspected of, or already suffering from such a disorder in an amount sufficient to cure, or at least partially arrest, the symptoms of the disorder (biochemical, histologic and/or behavioral), including its complications and intermediate pathological phenotypes in development of the disorder.

Effective doses of the compositions of the present invention, for the treatment of a condition vary depending upon many different factors, including means of administration,

target site, physiological state of the subject, whether the subject is human or an animal, other medications administered, and whether treatment is prophylactic or therapeutic. Usually, the subject is a human but non-human mammals including transgenic mammals can also be treated.

For passive immunization with an antibody, the dosage ranges from about 0.0001 to 100 mg/kg, and more usually 0.01 to 20 mg/kg, of the host body weight. For example dosages can be 1 mg/kg body weight or 10 mg/kg body weight or within the range of 1-10 mg/kg, *e.g.*, at least 1 mg/kg. Subjects can be administered such doses daily, on alternative days, weekly or according to any other schedule determined by empirical analysis. An exemplary treatment entails administration in multiple dosages over a prolonged period, for example, of at least six months. Additional exemplary treatment regimes entail administration once per every two weeks or once a month or once every 3 to 6 months. Exemplary dosage schedules include 1-10 mg/kg or 15 mg/kg on consecutive days, 30 mg/kg on alternate days or 60 mg/kg weekly. In some methods, two or more monoclonal antibodies with different binding specificities are administered simultaneously, in which case the dosage of each antibody administered falls within the ranges indicated.

Antibody is usually administered on multiple occasions. Intervals between single dosages can be weekly, monthly or yearly. In some methods, dosage is adjusted to achieve a plasma antibody concentration of 1-1000 mg/ml and in some methods 25-300 μg/ml. Alternatively, antibody can be administered as a sustained release formulation, in which case less frequent administration is required. Dosage and frequency vary depending on the half-life of the antibody in the subject. In general, human antibodies show the longest half-life, followed by humanized antibodies, chimeric antibodies, and nonhuman antibodies, in descending order.

The dosage and frequency of administration can vary depending on whether the treatment is prophylactic or therapeutic. In prophylactic applications, compositions containing the present antibodies or a cocktail thereof are administered to a subject not already in the disease state to enhance the subject's resistance. Such an amount is defined to be a "prophylactic effective dose." In this use, the precise amounts again depend upon the subject's state of health and general immunity, but generally range from 0.1 to 25 mg per dose, especially 0.5 to 2.5 mg per dose. A relatively low dosage is administered at relatively infrequent intervals over a long period of time. Some subjects continue to receive treatment for the rest of their lives.

In therapeutic applications, a relatively high dosage (*e.g.*, from about 1 to 200 mg of antibody per dose, with dosages of from 5 to 25 mg being more commonly used) at relatively short intervals is sometimes required until progression of the disease is reduced or terminated, and preferably until the subject shows partial or complete amelioration of symptoms of disease. Thereafter, the patent can be administered a prophylactic regime.

Therapeutic agents can be administered by parenteral, topical, intravenous, oral, subcutaneous, intraarterial, intracranial, intraperitoneal, intranasal or intramuscular means for prophylactic and/or therapeutic treatment. The most typical route of administration of a protein drug is intravascular, subcutaneous, or intramuscular, although other routes can be effective. In some methods, agents are injected directly into a particular tissue where deposits have accumulated, for example intracranial injection. In some methods, antibodies are administered as a sustained release composition or device, such as a Medipad™ device. The protein drug can also be administered via the respiratory tract, *e.g.*, using a dry powder inhalation device.

Agents of the invention can optionally be administered in combination with other agents that are at least partly effective in treatment of immune disorders.

The pharmaceutical compositions of the invention include at least one antibody of the invention in a pharmaceutically acceptable carrier. A "pharmaceutically acceptable carrier" refers to at least one component of a pharmaceutical preparation that is normally used for administration of active ingredients. As such, a carrier may contain any pharmaceutical excipient used in the art and any form of vehicle for administration. The compositions may be, for example, injectable solutions, aqueous suspensions or solutions, non-aqueous suspensions or solutions, solid and liquid oral formulations, salves, gels, ointments, intradermal patches, creams, lotions, tablets, capsules, sustained release formulations, and the like. Additional excipients may include, for example, colorants, taste-masking agents, solubility aids, suspension agents, compressing agents, enteric coatings, sustained release aids, and the like.

Agents of the invention are often administered as pharmaceutical compositions including an active therapeutic agent and a variety of other pharmaceutically acceptable components. See Remington's Pharmaceutical Science (15th ed., Mack Publishing Company, Easton, Pennsylvania (1980)). The preferred form depends on the intended mode of administration and therapeutic application. The compositions can also include, depending on the formulation desired, pharmaceutically acceptable, non-toxic carriers or diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, physiological phosphate-buffered saline, Ringer's solutions, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation may also include other carriers, adjuvants, or nontoxic, nontherapeutic, nonimmunogenic stabilizers and the like.

Antibodies can be administered in the form of a depot injection or implant preparation, which can be formulated in such a manner as to permit a sustained release of the active ingredient. An exemplary composition comprises monoclonal antibody at 5 mg/ml, formulated in aqueous buffer consisting of 50 mM L-histidine, 150 mM NaCl,

adjusted to pH 6.0 with HCl. Another example of a suitable formulation buffer for monoclonal antibodies contains 20 mM sodium citrate, pH 6.0, 10% sucrose, 0.1% Tween 80.

Typically, compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. The preparation also can be emulsified or encapsulated in liposomes or microparticles such as polylactide, polyglycolide, or copolymer for enhanced adjuvant effect, as discussed above (see Langer, Science 249: 1527 (1990) and Hanes, Advanced Drug Delivery Reviews 28:97 (1997)).

Therapies

Treatment of a subject suffering from a disease or disorder can be monitored using standard methods. Some methods entail determining a baseline value, for example, of an antibody level or profile in a subject, before administering a dosage of agent, and comparing this with a value for the profile or level after treatment. A significant increase (*i.e.*, greater than the typical margin of experimental error in repeat measurements of the same sample, expressed as one standard deviation from the mean of such measurements) in value of the level or profile signals a positive treatment outcome (*i.e.*, that administration of the agent has achieved a desired response). If the value for immune response does not change significantly, or decreases, a negative treatment outcome is indicated.

In other methods, a control value (*i.e.*, a mean and standard deviation) of level or profile is determined for a control population. Typically the individuals in the control population have not received prior treatment. Measured values of the level or profile in a subject after administering a therapeutic agent are then compared with the control value. A significant increase relative to the control value (*e.g.*, greater than one standard deviation from the mean) signals a positive or sufficient treatment outcome. A lack of significant increase or a decrease signals a negative or insufficient treatment outcome. Administration of agent is generally continued while the level is increasing relative to the control value. As before, attainment of a plateau relative to control values is an indicator that the administration of treatment can be discontinued or reduced in dosage and/or frequency.

In other methods, a control value of the level or profile (*e.g.*, a mean and standard deviation) is determined from a control population of individuals who have undergone treatment with a therapeutic agent and whose levels or profiles have plateaued in response to treatment. Measured values of levels or profiles in a subject are compared with the control value. If the measured level in a subject is not significantly different (*e.g.*, more than one standard deviation) from the control value, treatment can be discontinued. If the level in a subject is significantly below the control value, continued administration of

agent is warranted. If the level in the subject persists below the control value, then a change in treatment may be indicated.

In other methods, a subject who is not presently receiving treatment but has undergone a previous course of treatment is monitored for antibody levels or profiles to determine whether a resumption of treatment is required. The measured level or profile in the subject can be compared with a value previously achieved in the subject after a previous course of treatment. A significant decrease relative to the previous measurement (*i.e.*, greater than a typical margin of error in repeat measurements of the same sample) is an indication that treatment can be resumed. Alternatively, the value measured in a subject can be compared with a control value (mean plus standard deviation) determined in a population of subjects after undergoing a course of treatment. Alternatively, the measured value in a subject can be compared with a control value in populations of prophylactically treated subjects who remain free of symptoms of disease, or populations of therapeutically treated subjects who show amelioration of disease characteristics. In all of these cases, a significant decrease relative to the control level (*i.e.*, more than a standard deviation) is an indicator that treatment should be resumed in a subject.

The antibody profile following administration typically shows an immediate peak in antibody concentration followed by an exponential decay. Without a further dosage, the decay approaches pretreatment levels within a period of days to months depending on the half-life of the antibody administered. For example the half-life of some human antibodies is of the order of 20 days.

In some methods, a baseline measurement of antibody to a given antigen in the subject is made before administration, a second measurement is made soon thereafter to determine the peak antibody level, and one or more further measurements are made at intervals to monitor decay of antibody levels. When the level of antibody has declined to baseline or a predetermined percentage of the peak less baseline (*e.g.*, 50%, 25% or 10%), administration of a further dosage of antibody is administered. In some methods, peak or subsequent measured levels less background are compared with reference levels previously determined to constitute a beneficial prophylactic or therapeutic treatment regime in other subjects. If the measured antibody level is significantly less than a reference level (*e.g.*, less than the mean minus one standard deviation of the reference value in population of subjects benefiting from treatment) administration of an additional dosage of antibody is indicated.

Additional methods include monitoring, over the course of treatment, any art-recognized physiologic symptom (*e.g.*, physical or mental symptom) routinely relied on by researchers or physicians to diagnose or monitor disorders.

The following examples are included for purposes of illustration and should not be construed as limiting the invention.

Exemplification

Throughout the examples, the following materials and methods were used unless otherwise stated.

Materials and Methods

In general, the practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, recombinant DNA technology, immunology (especially, *e.g.*, antibody technology), and standard techniques in electrophoresis. See, *e.g.*, Sambrook, Fritsch and Maniatis, Molecular Cloning: Cold Spring Harbor Laboratory Press (1989); Antibody Engineering Protocols (Methods in Molecular Biology), 510, Paul, S., Humana Pr (1996); Antibody Engineering: A Practical Approach (Practical Approach Series, 169), McCafferty, Ed., Irl Pr (1996); Antibodies: A Laboratory Manual, Harlow *et al.*, C.S.H.L. Press, Pub. (1999); and Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley & Sons (1992).

Generation of Antibodies and Antigen-Binding Fragments Thereof

The selection, cloning, and manufacture of antibodies, for example, chimeric, humanized, monoclonal, and single-chain antibodies is well described in the art. In addition, the humanization of hu5c8 mAb has been described previously. See Lederman, 1992 and Karpusas, 2001, respectively. This antibody is available from the ATCC (PTA-4931). The 5c8 antibody was stably expressed in NS0 myeloma cells and purified by Protein A and gel filtration chromatography. SDS-PAGE and analytical gel filtration chromatography demonstrated that the protein formed the expected disulfide linked tetramer. The single-chain antibodies of the invention were typically expressed in *E. coli* and immunopurified using standard techniques.

AQC2 scFv production

AQC2 scFv is expressed by plasmid pKJS217. This plasmid contains 318 nucleotides of the AQC2 light chain encoding the 106 amino acid light chain variable region followed in frame by 45 nucleotides encoding 3 copies of a GGGGS linker moiety. The linker is followed in frame by 360 nucleotides encoding the 120 amino acid AQC2 heavy chain variable region. Immediately following the heavy chain variable region is an enterokinase cleavage site and myc and HIS tags. Expression was done in *E.coli* and is driven by the ara-BAD promoter and the protein is directed to the periplasmic space by an 80 nucleotide fragment encoding the g111 peptide from the bacteriophage fd. This peptide was cleaved from the protein during periplasmic export.

5C8 Fab production

The 5C8 Fab fragment was expressed by the bicistronic plasmid pBEF064. The first cistron contains 354 nucleotides of the 5C8 heavy chain encoding the 118 amino acid heavy chain variable region followed in frame by 306 nucleotides encoding the first 102 amino acids of the human IgG1 constant domain and 18 nucleotides encoding a 6 histidine tag. A second ribosome entry site is located 7 nucleotides after the end of the heavy chain cistron. The second cistron contains 333 nucleotides encoding the 111 amino acid 5C8 light chain variable region followed in frame by 321 nucleotides encoding the 107 amino acid light chain constant domain. Expression was done in *E.coli* and is driven by the ara-BAD promoter and the heavy and light chains are directed to the periplasmic space by the OmpA (heavy chain) and PhoA (light chain) periplasmic localization signals. The periplasmic localization signals are cleaved from the protein during periplasmic export.

Binding Assays

Binding assays were typically performed using the KinExA™ kit. The assay is carried out by passing a dilute solution of the antibody (or antigen-binding fragment) through the column provided in the kit, and some of the antibody (or the antigen-binding fragment thereof) interacts with the antigen on the bead. The antibody (or the fragment) is then detected with a secondary anti-human IgG heavy and light chain antibody conjugated with the fluorescent dye Cy5 (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA). The concentration of the antibody (or the fragment) is set so that the signal from the fluorescent dye is proportional to the concentration of protein. To obtain the solution phase affinity of the interaction, the antibody (or the fragment) is mixed with a dilution series of soluble antigen. These proteins (antibody and antigen) are allowed to reach equilibrium during a three-hour incubation at room temperature or an overnight incubation at 4°C. The mixture is flowed over the antigen-containing column, and the signal is proportional to the amount of unbound antibody (or antibody fragment) that remains in solution. The resulting data can be plotted on a linear-log scale graph and fit to a quadratic curve by non-linear regression, which gives a value for the K_D.

Binding assay 5C8-CD40L

An ELISA-based competitive binding assay was done. Anti c-myc mAb was coated onto NUNC Maxisorb plates at 10 ug/mL in PBS for 2 hrs at room temperature. Serial dilutions of unlabeled 5C8 Fab (mutants or wildtype) were made and mixed with equal volumes of fixed concentration (30 ng/ml) of biotin-labeled 5C8 Fab competitor, and added to the plate. After 2 hours incubation at room temperature, the plate was

washed and bound biotin-labeled 5C8 Fab competitor was detected with streptavidin-HRP. Binding affinities were obtained from four parameter curve fits.

Computer Modeling Metrics and Formulae

For carrying out the optimization of an antibody according to the invention, the following metrics and formulae can be used. For example, the free energy of binding difference between the electrostatic free energy in the bound and the unbound state of antibody can be represented as such, $\Delta G_{binding}=G^{bound}-G^{unbound}$ (see FIG. *1a)*. Because the dielectric model includes responses that affect the entropy as well as the enthalpy, the electrostatic energy is considered to be a free energy. The free energy of each state is expressed as a sum of coulombic and reaction-field (hydration) terms involving the antigen (L), the antibody (or antigen-binding fragment thereof) (R), and their interaction (L-R):

$$G^{state}=G_{coul,L}^{state}+G_{coul,R}^{state}+G_{coul,L-R}^{state}+G_{hyd,L}^{state}+G_{hyd,R}^{state}+G_{hyd,L-R}^{state} \quad (1)$$

This results in the following expression for the binding free energy,

$$\Delta G_{binding}=\Delta G_{coul,L-R}+\Delta G_{hyd,L-R}+\Delta G_{hyd,L}+\Delta G_{hyd,R} \quad (2)$$

where the fact that the geometry of point charges in the antigen and antibody remain fixed is used in the model to cancel the coulombic self contribution of antibody and antigen and where the two L-R terms are due only to the bound state because the antibody and antigen are assumed not to interact in the unbound state. (Note, however, that the charge distribution for the antigen need not be the same in the bound and unbound states. If they are different, this adds a constant to $\Delta G_{binding}$ that can be dropped in defining ΔG_{var} in Eq. (3)). Thus, Eq. (2) describes the electrostatic binding free energy as a sum of desolvation contributions of the antibody and the antigen (which are unfavorable) and solvent screened electrostatic interaction in the bound state (which is usually favorable). Since the goal is to vary the antibody charge distribution to optimize the electrostatic binding free energy and the last term simply adds a constant, a relevant variational binding energy is defined,

$$\Delta G_{var}=\Delta G_{coul,L-R}+\Delta G_{hyd,L} \quad (3)$$

in which the first two terms on the right hand side (RHS) of Eq. (2) have been combined into a screened interaction term and the constant term has been dropped. Note that

$$\Delta G_{coul,L-R}=\sum_{i\in R} q_i V_L^{bound}(\vec{r}_i)=\sum_{i\in R} q_i[V_{coul,L}^{bound}(\vec{r}_i)+V_{hyd,L}^{bound}(\vec{r}_i)] \quad (4)$$

and

$$\Delta G_{hyd,L}=\frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{bound}(\vec{r}_i)-\frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{unbound}(\vec{r}_i) \quad (5)$$

where V_L^{state} is the total electrostatic potential in the indicated state due to the antibody charge distribution only and $V_{term,L}^{state}$ is the coulombic or reaction-field (hydration) term, as indicated. The summations are over atomic point charges in the antibody ($i \in L$) or antigen $j \in R$). The factor of 1/2 in Eq. (5) is due to the fact that the antibody charge distribution interacts with the self-induced reaction field. $V_{coul,L}^{bound}$, $V_{hyd,L}^{bound}$, and $V_{hyd,L}^{unbound}$, the three electrostatic potentials in Eqs. (4) and (5), are expressed in terms of the given geometry and charge distribution by solving the boundary-value problem shown in FIG. *1b*. A charge distribution (corresponding to the antibody) is embedded in a sphere of radius R. The center of the sphere is taken as the origin of coordinates (unprimed) but the charge distribution in multipoles is expanded about a second origin (primed) translated a distance d along the z-axis, so that

$$\vec{r}(r,\theta,\phi) = \vec{d}(d,\theta'=0,\phi'=0) + \vec{r}\,'(r',\theta',\phi'). \qquad (6)$$

The potential everywhere satisfies the Poisson equation. Inside the sphere, it may be written as,

$$V_{in}(\vec{r}) = \sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r}_i|} + \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{l,m} r^l Y_{l,m}(\theta, \phi) \qquad (7)$$

where the first term on the RHS is the coulombic and the second is the reaction-field (hydration) potential, and the summation over i corresponds to the antibody point charges. Outside the sphere, the coulombic and reaction-field potential can be combined and written as,

$$V_{out}(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{B_{l,m}}{r^{l+1}} Y_{l,m}(\theta, \phi) \qquad (8)$$

where $A_{l,m}$ and $B_{l,m}$ are to be determined by the proper boundary conditions and $Y_{l,m}(\theta, \phi)$ are the spherical harmonics. The coulombic term in Eq. (7) is expanded in spherical harmonics and multipoles of the charge distribution about the center of the sphere. Here the origin of the multipole expansion is shifted to $\vec{d}$,

$$\sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r}_i|} = \sum_i \frac{q_i}{\epsilon_1 |(\vec{r} - \vec{d}) - \vec{r}_i'|} = \sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r}_i'|} \qquad (9)$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} Q'^*_{l,m} \frac{Y_{l,m}(\theta', \phi')}{\epsilon_1 r^{l+1}} \qquad (10)$$

where $Q'_{l,m}$ is a spherical multipole expanded about the primed origin, $\vec{d}$,

$$Q'^*_{l,m} \equiv \sum_i q_i r_i'^l Y^*_{l,m}(\theta_i', \phi_i'). \qquad (11)$$

The definition of the $Y_{l,m}(\theta, \phi)$ used by Jackson is adopted (J, D. Jackson, *Electrodynamics*, 2nd ed, (John Wiley and Sons. New York. 1975).

The expression in Eq. (10) is valid for r'>r'i (*i.e.*, outside the antibody or, more precisely, outside the sphere whose center is at $\vec{d}$ and whose radius is the longest distance between $\vec{d}$ and a point charge). To substitute into Eq. (7) and combine terms involving spherical harmonics, first $Y_{l,m}(\theta', \phi')/r'^{l+1}$ of Eq. (10) was expanded in terms of $Y_{l,m}(\theta, \phi)/r^{l+1}$. This was done using the results of Greengard (*The Rapid Evaluation* of *Potential Fields in Particle Systems* MIT Press, Cambridge, Mass., 1988) which state that for r>d,

$$\frac{Y_{l,m}(\theta', \phi')}{r'^{l+1}} = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} K_{l',m',l,m} \left[\frac{4\pi(2l+1)}{(2l'+1)(2l'+2l+1)} \right]^{\frac{1}{2}} \tag{12}$$

$$d^{l'} Y^{*}_{l',m'}(\theta_d, \phi_d) \frac{Y_{l'+l,m'+m}(\theta, \phi)}{r^{l'+l+1}}$$

where

$$K_{l',m',l,m} = \left[\frac{(l'+l+m'+m)!(l'+l-m'-m)!}{(l'+m')!(l'-m')!(l+m)!(l-m)!} \right]^{\frac{1}{2}}. \tag{13}$$

Since a geometry with θ_d=0 has been used, only m'=0 terms in Eq. (12) are non-vanishing, in which case Eq. (10) becomes,

$$\sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r}_i|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{\epsilon_1} \left(\frac{4\pi}{2l+1} \right)^{\frac{1}{2}} \tag{14}$$

$$Q'^{*}_{l,m} \sum_{l'=0}^{\infty} K_{l',0,l,m} d^{l'} \left(\frac{4\pi}{2l'+2l+1} \right)^{\frac{1}{2}} \frac{Y_{l'+l,m}(\theta, \phi)}{r^{l'+l+1}}$$

in which the multipole distribution was taken about the point $\vec{d}$, but the potential was expressed as a summation of spherical harmonics about the large-sphere center. The above equation can also be written as,

$$\sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r}_i|} = \tag{15}$$

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{\epsilon_1} \left(\frac{4\pi}{2l+1} \right)^{\frac{1}{2}} \frac{Y_{l,m}(\theta, \phi)}{r^{l+1}} \sum_{l'=|m|}^{l} K_{l-l',0,l',m} d^{l-l'} \left(\frac{4\pi}{2l'+1} \right)^{\frac{1}{2}} Q'^{*}_{l',m}$$

where terms with the same $Y_{l,m}(\theta, \phi)$ are grouped together, as opposed to Eq. (14), where terms with the same $Q'^{*}_{l,m}$ are grouped.

Upon substituting Eq. (15) into Eq. (7) and matching boundary conditions at r=R or room temperature,

$$V_{in}|_{r=R} = V_{out}|_{r=R} \tag{16}$$

$$\epsilon_1 \frac{\partial V_{in}}{\partial r}\bigg|_{r=R} = \epsilon_2 \frac{\partial V_{out}}{\partial r}\bigg|_{r=R} \tag{17}$$

the hydration (reaction-field) potential inside the sphere is,

$$V_{hyd}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{l,m} r^l Y_{l,m}(\theta, \phi) \tag{18}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} r^l Y_{l,m}(\theta, \phi)\left(\frac{C_l}{R^{2l+1}}\right)$$

$$\sum_{l'=|m|}^{l} K_{l-l',0,l',m} d^{l-l'}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} Q'^{*}_{l',m} \tag{19}$$

where

$$C_l = \frac{(\epsilon_1 - \epsilon_2)}{\epsilon_1 [\epsilon_2 + l\epsilon_1/(l+1)]}. \tag{20}$$

The various V's can be rewritten, with their dependence on the $Q'^{*}_{l,m}$ made explicit. $V'_{coul,L}{}^{bound}$ is given by Eq. (20), $V_{hyd,L}{}^{bound}$ is given by Eq. (19) but rewritten so that the terms with the same $Q'^{*}_{l,m}$ are collected, and $V_{hyd,L}{}^{unbound}$ is given by Eq. (19) with R=a and d=0.

$$V_{coul,L}^{bound}(\vec{r}') = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{2l+1} Q'^{*}_{l,m} \frac{Y_{l,m}(\theta', \phi')}{\epsilon_1 r'^{l+1}} \tag{21}$$

$$V_{hyd,L}^{bound}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=l}^{\infty}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} \tag{22}$$

$$\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} Q'^{*}_{l,m} K_{l'-l,0,l,m} d^{l'-l}\left(\frac{C_{l'}}{R^{2l'+1}}\right) r^{l'} Y_{l',m}(\theta, \phi)$$

$$V_{hyd,L}^{unbound}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{2l+1}\left(\frac{C_l}{a^{2l+1}}\right) Q'^{*}_{l,m} r^l Y_{l,m}(\theta, \phi) \tag{23}$$

Substituting into Eq. (4), the dependence of $\Delta G_{int,L-R}$ on the $Q'^{*}_{l,m}$ is made explicit,

$$\Delta G_{int,L-R} = \sum_{j\in R} q_j \left[V_{coul,L}^{bound}(\vec{r}_j) + V_{hyd,L}^{bound}(\vec{r}_j)\right] \tag{24}$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Q'^{*}_{l,m} \sum_{j \in R} q_j \left[\left(\frac{4\pi}{2l+1} \right) \frac{Y_{l,m}(\theta'_j, \phi'_j)}{\epsilon_1 r'^{l+1}_j} + \right. \tag{25}$$

$$\sum_{l'=l}^{\infty} \left(\frac{4\pi}{2l+1} \right)^{\frac{1}{2}} \left(\frac{4\pi}{2l'+1} \right)^{\frac{1}{2}} K_{l'-l,0,l,m} d^{l'-l} \left(\frac{C_{l'}}{R^{2l'+1}} \right)$$

$$\left. r_j^{l'} Y_{l',m}(\theta_j, \phi_j) \right]$$

$$\equiv \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m} Q'^{*}_{l,m} \tag{26}$$

where in the last line the element $\alpha_{l,m}$ is defined, which is independent of the $Q'^{*}_{l,m}$, to be the factor multiplying $Q'^{*}_{l,m}$ in Eq. (25). Each $\alpha_{l,m}$ expresses the contribution of a multipole to $\Delta G_{int,L\text{-}R}$ and contains all information concerning the antigen charge distribution required to obtain ΔG_{var}. For $\Delta G_{hyd,L}$ it is useful to re-express Eq. (5) in terms of the $Q'^{*}_{l,m}$, the multipoles describing the antibody charge distribution, rather than the individual charges, q_i.

$V(\vec{r})$ is expanded around the center of the multipole expansion, $\vec{d}$,

$$\sum_{i \in L} q_i V(\vec{r}_i) = \sum_{i \in L} q_i V\left(\vec{d} + \vec{r}_i\right) \tag{27}$$

$$= \sum_{i \in L} q_i \left[V(\vec{d}) + \vec{r}_i \cdot \vec{\nabla} V(\vec{d}) + \ldots \right]. \tag{28}$$

It has been shown by Rose (M. E. Rose, J. Math. & Phys. 37, 215 (1958); M. E. Rose, *Elementary Theory* of *Angular Momentum* (John Wiley and Sons, New York, 1957)) that in spherical coordinates the expansion becomes,

$$\sum_{i \in L} q_i V\left(\vec{d} + \vec{r}_i\right) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{(2l+1)!!} Q'^{*}_{l,m} y_{l,m}(\vec{\nabla}) V(\vec{d}) \tag{29}$$

where

$$y_{l,m}(\vec{r}) \equiv r^l Y_{l,m}(\theta,\phi) \tag{30}$$

and $y_{l,m}(\vec{\nabla})$ is the operator obtained by replacing $\vec{r}$ with $\vec{\nabla}$.

For positive m and when $y_{l,m}(\vec{\nabla})$ operates on a solution of the Laplace equation (*i.e.* $r^l Y_{l,m}(\theta, \phi)$ or $Y_{l,m}(\theta, \phi)/r^{l+1}$), it has been shown that,

$$y_{L,n}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^n}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_1^m \nabla_0^{l-m} \quad (31)$$

for $m \geq 0$.

The double factorial is defined as

$$(2l+1)!! = (2l+1)\cdot(2l-1)\cdot(2l-3)\ldots 3\cdot 1 \quad (32)$$

$$= \frac{(2l+1)!}{2^l l!} \quad (33)$$

and the spherical partial derivatives are

$$\nabla_1 = -\frac{1}{\sqrt{2}}(\nabla_x + i\nabla_y),\ \nabla_{-1} = \frac{1}{\sqrt{2}}(\nabla_x - i\nabla_y),\ \nabla_0 = \nabla_z. \quad (34)$$

To compute $y_{l,m}(\vec{\nabla})$ for negative m, the fact that $Y_{l,-m}(\theta, \phi)=(-1)^m Y^*_{l,m}(\theta, \phi)$ is used and the definitions of spherical partial derivatives in Eq. (34) to obtain,

$$y_{l,-m}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_{-1}^m \nabla_0^{l-m} \quad (35)$$

for $m \geq 0$.

The hydration energy of the bound antibody is then

$$G_{hyd,L}^{bound} = \frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{bound}\left(\vec{d} + \vec{r}_i\right) = \frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \frac{4\pi}{(2l'+1)!!} Q_{l',m'}^{l'} y_{l',m'}(\vec{\nabla}) V_{hyd,L}^{bound}(\vec{d}) \quad (36)$$

$$= \frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \frac{4\pi}{(2l'+1)!!} Q_{l',m'}^{l'} y_{l',m'} \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l''=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l''+1}\right)^{\frac{1}{2}} \times$$

$$Q_{l,m}^{*} K_{l''-l,0,l,m} d^{l''-l}\left(\frac{C_{l''}}{R^{2l''+1}}\right) y_{l',m'}(\vec{\nabla})\left(r^{l''} Y_{l'',m}(\theta, \phi)\right)\Big|_{\vec{r}=\vec{d}}. \quad (37)$$

To evaluate $y_{l,m}(\vec{\nabla})$ in Eq. (37), Eq. (31) and the gradient formula are used (M. E. Rose, *Elementary Theory* of *Angular Momentum* (John Wiley and Sons, New York, 1957))

$$\vec{\nabla}(\Phi(r)Y_{l,m}(\theta, \phi)) = -\left(\frac{l+1}{2l+1}\right)^{\frac{1}{2}}\left(\frac{d\Phi(r)}{dr} - \frac{l}{r}\Phi(r)\right)T_{l,l+1,m}(\theta, \phi) + \quad (38)$$

$$\left(\frac{l}{2l+1}\right)^{\frac{1}{2}}\left(\frac{d\Phi(r)}{dr} + \frac{l+1}{r}\Phi(r)\right)T_{l,l-1,m}(\theta, \phi)$$

where

$$T_{l,l',m}(\theta,\phi) \equiv \sum_{m'\in\{-1,0,1\}} C(l',1,l;m-m',m')Y_{l',m-m'}(\theta,\phi)\hat{\xi}_{m'} \qquad (39)$$

the $C(l', 1, l; m-m', m')$ are the vector addition (or Clebsch-Gordon) coefficients frequently encountered in the study of angular momentum shown in Table 1 (of Rose), and $\hat{\xi}_{m'}$

are spherical unit vectors,

$$\hat{\xi}_1 = -\frac{1}{\sqrt{2}}(\hat{x}+i\hat{y}),\ \hat{\xi}_{-1} = \frac{1}{\sqrt{2}}(\hat{x}-i\hat{y}),\ \hat{\xi}_0 = \hat{z}. \qquad (40)$$

Accordingly,

$$\vec{\nabla} = \hat{x}\nabla_x + \hat{y}\nabla_y + \hat{z}\nabla_z = -\hat{\xi}_1\nabla_{-1} - \hat{\xi}_{-1}\nabla_1 + \hat{\xi}_0\nabla_0. \qquad (41)$$

From Eqs. (38) through (41),

$$\nabla_u(r^l Y_{l,m}(\theta,\phi) = (-1)^u[l(2l+1)]^{1/2}C(l-1,1,l;m+u,-u)r^{l-1}Y_{l-1,m+u}(\theta,\phi)$$

Using Table I, Eq. (31), and Eqs. (37) through (42), the following intermediate results are obtained,

$$\nabla_0^{l'-m'}\left(r^{l''}Y_{l'',m}\right) = \qquad (43)$$

$$\left[\frac{(2l''+1)(l''+m)!(l''-m)!}{(2l''-2l'+2m'+1)(l''-m-l'+m')!\,(l''+m-l'+m')!}\right]^{\frac{1}{2}} r^{l''-l'+m'}Y_{l''-l'+m',m}.$$

$$\nabla_1^{m'}\left(r^{l''-l'+m'}Y_{l''-l'+m',m}\right) = \qquad (44)$$

$$(-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''-m-l'-m')!}{2^{m'}(2l''-2l'+1)(l''-m-l'-m')!}\right]^{\frac{1}{2}} r^{l''-l'}Y_{l''-l',m+m'}$$

$$\nabla_{-1}^{m'}\left(r^{l''-l'+m'}Y_{l''-l'+m',m}\right) = \qquad (45)$$

$$(-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''+m-l'+m')!}{2^{m'}(2l''-2l'+1)(l''+m-l'-m')!}\right]^{\frac{1}{2}} r^{l''-l'}Y_{l''-l',m-m'}$$

and the final expression for the hydration energy of the antibody in the bound state,

$$G_{hyd,L}^{bound} = \frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{bound}(\vec{r}_i) = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty} Q_{l,m}^{l*}Q_{l',m}^{l'} \sum_{l''=\max(l,l')}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}}\frac{C_{l''}}{R^{2l''+1}}\times$$

$$\frac{(l''+m)!(l''-m)!}{(l''-l)!(l''-l')!}\left[\frac{1}{(l+m)!(l-m)!(l'+m)!(l'-m)!}\right]^{\frac{1}{2}} d^{2l''-l-l'} \qquad (46)$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \beta_{l,m,l',m'}Q_{l,m}^{l*}Q_{l',m'}^{l'} \qquad (47)$$

where $\beta_{l,m,l',m'}$ is defined by the above two equations; note that $\beta_{l,m,l',m'}$ is zero for m' m.

The hydration energy of the unbound antibody is obtained by setting d=0 and R=α in Eq. (46),

$$G_{hyd,L}^{unbound} = \frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{unbound}(\vec{r}_i) = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\left(\frac{C_l}{a^{2l+1}}\right)Q_{l,m}^{\prime *}Q_{l,m}^{\prime} \quad (48)$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q_{l,m}^{\prime *}Q_{l,m}^{\prime} \quad (49)$$

where $\gamma_{l,m}$ is defined by Eqs. (48) and (49). Then, $\gamma_{l,m}$ is written as a function of both *l* and *m* for notational convenience, although there is no formal dependence on *m*.

Thus ΔG_{var} has been expressed as a function of the multipoles of the antibody charge distribution, $Q'_{l,m}$ (expanded about the center of the antibody sphere) and the elements $\alpha_{l,m}$, $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ which do not depend on $Q'_{l,m}$. Combining Eqs. (26), (47) and (49) gives

$$\Delta G_{var} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q_{l,m}^{\prime *} + \quad (50)$$

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\beta_{l,m,l',m'}Q_{l,m}^{\prime *}Q_{l',m'}^{\prime} - \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q_{l,m}^{\prime *}Q_{l,m}^{\prime}.$$

Note that only the $\alpha_{l,m}$ depend on the antigen charges, while the $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ depend solely on the geometry of the bound and unbound states. While ΔG_{var}^{opt} is a real quantity, the $\alpha_{l,m}$ and $Q'_{l,m}$ are complex and the products $\alpha_{l,m}Q'^{*}_{l,m}$ and $Q'^{*}_{l,m}Q'_{l',m}$ involve summations over terms of the form $Y^{*}_{l',m}(\theta', \phi')Y_{l,m}(\theta, \phi)$; note that the $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ are real. Then ΔG_{var}^{opt} is rewritten in terms of the real and imaginary parts of $\alpha_{l,m}$ and $Q'_{l,m}$,

$$\Delta G_{var} = \sum_{l=0}^{\infty}\left[\alpha_{l,0}Q_{l,0}^{\prime} + 2\sum_{m=1}^{l}(\mathrm{Re}\alpha_{l,m}\mathrm{Re}Q_{l,m}^{\prime} + \mathrm{Im}\alpha_{l,m}\mathrm{Im}Q_{l,m}^{\prime})\right] + \quad (51)$$

$$\sum_{l=0}^{\infty}\sum_{l'=0}^{\infty}\left[\beta_{l,0,l',0}Q_{l,0}^{\prime}Q_{l',0}^{\prime} + \right.$$

$$\left. 2\sum_{m=1}^{l}\beta_{l,m,l',m}(\mathrm{Re}Q_{l,m}^{\prime}\mathrm{Re}Q_{l',m}^{\prime} + \mathrm{Im}Q_{l,m}^{\prime}\mathrm{Im}Q_{l',m}^{\prime})\right] -$$

$$\sum_{l=0}^{\infty}\left[\gamma_{l,0}Q_{l,0}^{\prime 2} + 2\sum_{m=1}^{l}\gamma_{l,m}(\mathrm{Re}Q_{l,m}^{\prime 2} + \mathrm{Im}Q_{l,m}^{\prime 2})\right]$$

(where summations over m are excluded for *l*=0) by noting again that $Y_{l,-m}(\theta, \phi)=(-1)^{m}Y^{*}_{l,m}(\theta, \phi)$ and

$$Y^*_{l,m}(\theta',\phi')Y_{l,m}(\theta,\phi)+Y^*_{l,-m}(\theta',\phi')Y_{l,-m}(\theta,\phi)=Y^*_{l,m}(\theta',\phi')Y_{l,m}(\theta,\phi)$$
$$+Y_{l,m}(\theta',\phi')Y^*_{l,m}(\theta,\phi) \quad (52)$$
$$=2[\mathrm{Re}Y_{l,m}(\theta',\phi')\cdot\mathrm{Re}Y_{l,m}(\theta,\phi)+\mathrm{Im}Y_{l,m}(\theta',\phi')\cdot\mathrm{Im}Y_{l,m}(\theta,\phi)]. \quad (53)$$

The new variables $\mathrm{ReQ'}_{l,m}$ and $\mathrm{ImQ'}_{l,m}$ are re-indexed and renamed Q_i as follows,

$$\{Q'_{0,0},Q'_{1,0},\mathrm{ReQ'}_{1,1},\mathrm{ImQ'}_{1,1},Q'_{2,0},\mathrm{ReQ'}_{2,1},\mathrm{ImQ'}_{2,1},$$
$$\mathrm{ReQ'}_{2,2},\ldots\}\leftrightarrow\{Q_1,Q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8,\ldots\}. \quad (54)$$

and similar transformations are used to create α_i, β_{ij}, and γ_i. Eq. (51) can then be written as

$$\Delta G_{\mathrm{var}} = \sum_{i=1}^{\infty}\alpha_i Q_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\beta_{ij}Q_iQ_j - \sum_{i=1}^{\infty}\gamma_i Q_i^2 \quad (55)$$

$$= \sum_{i=1}^{\infty}\alpha_i Q_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(\beta_{ij}-\delta_{ij}\gamma_i)Q_iQ_j \quad (56)$$

or in matrix notation,

$$\Delta G_{\mathrm{var}} = \vec{Q}^T\overleftrightarrow{B}\,\vec{Q}+\vec{Q}^T\vec{A} \quad (57)$$

$$=\left(\vec{Q}+\frac{1}{2}\overleftrightarrow{B}^{-1}\vec{A}\right)^T\overleftrightarrow{B}\left(\vec{Q}+\frac{1}{2}\overleftrightarrow{B}^{-1}\vec{A}\right)-\frac{1}{4}\vec{A}^T\overleftrightarrow{B}^{-1}\vec{A} \quad (58)$$

where $\vec{Q}$ is the vector formed by the Q_i, $\vec{A}$ is the vector formed by the α_i, $\overleftrightarrow{B}$ is the symmetric matrix formed by the $(\beta_{ij}-\delta_{ij}\gamma_i)$, and completion of the square has been used to arrive at Eq. (58). Since $\vec{Q}^T\overleftrightarrow{B}\,\vec{Q}$ in Eq. (57) corresponds to the antibody desolvation penalty, which must be greater than zero for chemically reasonable geometries, the matrix $\overleftrightarrow{B}$ is positive definite and the extreme of ΔG_{var} is a minimum (G. Strang, *Introduction to Applied Mathematics* (Wellesley-Cambridge Press, Wellesley, Mass., 1986).

From Eq. (58) the optimum values of the multipoles, $\vec{Q}^{opt}$ and the minimum variational binding energy, $\Delta G_{var}{}^{opt}$ are obtained,

$$\vec{Q}^{opt} = -\frac{1}{2}\overleftrightarrow{B}^{-1}\vec{A} \quad (59)$$

$$\Delta G_{\mathrm{var}}^{opt} = -\frac{1}{4}\vec{A}^T\overleftrightarrow{B}^{-1}\vec{A}. \quad (60)$$

$\Delta G_{var}{}^{opt}$ is always negative because $\overleftrightarrow{B}^{-1}$ is also positive definite.

To solve for the optimal multipole distribution with the monopole (total charge) fixed ($Q_1=$ ~~Q~~), the equation for the remaining optimal multipoles ($i\neq1$) is,

$$2\sum_{j\neq 1}(\beta_{ij}-\delta_{ij}\gamma_i)Q_j^{opt}+(2\beta_{i1}Q+\alpha_i)=0 \qquad (61)$$

which is analogous to Eq. (59).

The above matrix equations, with the dimension truncated at $i_{max}=(l_{max}+1)^2$, can be solved numerically by relatively modest computational resources. In practice, since the α_i and β_{ij}, contain a summation over an infinite number of terms, a second cutoff value of l_{cut} must be used to truncate the innermost sum in Eqs. (25) and (46). When l_{max} and l_{cut} are sufficiently large, ΔG_{var}^{opt} converges and the incremental advantage of including more multipoles essentially vanishes.

For any given antigen and geometry, the present description has thus described a method to determine the charge distribution of the tightest binding antibody as a set of multipoles. The deviation of the binding free energy from the optimum for any test antibody

can be calculated by subtracting Eq. (60) from Eq. (58) and using Eq. (59) to eliminate $\vec{A}$,

Table 1 – Vector Addition Coefficients

$C(l', 1, l; m - m', m')$.[a]

	$m' = 1$	$m' = 0$	$m' = -1$
$l = l' + 1$	$\left[\frac{(l'+m)(l'+m+1)}{(2l'+1)(2l'+2)}\right]^{1/2}$	$\left[\frac{(l'-m+1)(l'+m+1)}{(2l'+1)(l'+1)}\right]^{1/2}$	$\left[\frac{(l'-m)(l'-m+1)}{(2l'+1)(2l'+2)}\right]^{1/2}$
$l = l'$	$-\left[\frac{(l'+m)(l'-m+1)}{2l'(l'+1)}\right]^{1/2}$	$\frac{m}{[l'(l'+1)]^{1/2}}$	$\left[\frac{(l'-m)(l'+m+1)}{2l'(l'+1)}\right]^{1/2}$
$l = l' - 1$	$\left[\frac{(l'-m)(l'-m+1)}{2l'(2l'+1)}\right]^{1/2}$	$-\left[\frac{(l'-m)(l'+m)}{l'(2l'+1)}\right]^{1/2}$	$\left[\frac{(l'+m+1)(l'+m)}{2l'(2l'+1)}\right]^{1/2}$

[a]from reference 4

EXAMPLE 1
METHODS OF IMPROVING THE ANTIGEN-BINDING AFFINITY OF AN ANTI-INTEGRIN ANTIBODY

In this example, methods for improving the binding affinity of an antibody against a therapeutically relevant antigen target, are described.

As proof-of-principle, the method of the invention was applied to an antibody against VLA-1 integrin, a cell-surface receptor for collagen and laminin, and in particular, the monoclonal antibody AQC2, which was raised against the human VLA-1 receptor by affinity maturation in mice. AQC2 inhibits the pathological processes mediated by VLA-1 integrin (see, *e.g.*, WO 02/083854).

A variant of ACQ2 with two mutations binds to VLA-1 with 100-fold less affinity than the wild-type antibody. In an effort to restore this binding, electrostatic charge optimization techniques were applied to a crystal structure of the antibody-antigen complex in a two-level procedure to suggest improved-affinity mutants. First, electrostatic charge optimization was used to determine the position(s) of the CDR residue(s) that are sub-optimal for binding (Lee and Tidor, *J. Chem. Phys.* 106:8681-8690, 1997; Kangas and Tidor, *J. Chem. Phys.* 109:7522-7545, 1998). Second, a set of CDR mutations were then determined for further computational analysis. Based on these calculations, the binding affinity was determined for 36 modified antibodies having a single mutation (*i.e.*, 36 "single mutants") and 10 antibodies having two mutations (*i.e.*, ten "double mutants"). It was predicted that 26 of the single mutants would be electrostatically favorable relative to the wild-type antibody, and that 15 would bind better with a full energy function including a van der Waals energy term and a solvent accessible surface area term. These terms are unrelated to electrostatic forces, but they were calculated to ensure that the designed mutations did not contact other residues and would not reduce the amount of buried surface area significantly; increased buried surface area in complex formation is usually beneficial (see the "Full Energy" column of the table below). Additionally, it was predicted that many of the double mutants would be more favorable than the wild-type complex and that the effects would be partially additive with respect to the single mutants.

The mutation predictions can be categorized as involving (1) mutations at the interaction interface involving residues that become partially buried upon binding (interactions are improved by making hydrogen bonds with the antibody); (2) mutations of polar residues on the antibody that become buried upon binding and thus pay a desolvation penalty but do not make any direct electrostatic interactions with the antibody (improvements are usually made by mutation to a hydrophobic residue with similar shape to the wild-type residue or by adding a residue that can make favorable electrostatic interactions); and (3) mutations of surface residues on the antibody that are in regions of

uncomplementary potentials. These mutations are believed to improve long-range electrostatic interactions between the antibody and antigen without perturbing packing interactions at the binding interface.

Based on results from a charge optimization, mutations were determined for computational analysis (the optimal charge distributions and design mutations that were closer to optimal than the current residue were examined; this process was done by inspection). A charge optimization gave charges at atom centers but did not yield actual mutation(s). A round of charge optimizations was performed with various constraints imposed to represent natural side chain characteristics. For example, an optimization was performed for a net side chain charge of –1, 0, and +1 with the additional constraint that no atom's charge exceeded an absolute value of 0.85 electron charge units.

The crystal structure of the VLA-1/AQC2 complex (PDB code: 1MHP) was prepared using standard procedures for adding hydrogens with the program CHARMM (Accelrys, Inc., San Diego, CA). N-acetamide and N-methylamide patches were applied to the N termini and C-termini, respectively. There was missing density for residues 288-293 in one of the complexes (Model 1), but no attempt was made to rebuild the density. Using a continuum electrostatics model, an electrostatic charge optimization was performed on each side chain of the amino acids in the CDRs of the ACQ2 antibody. Appropriate side chain mutations were then determined based on the potential gain in electrostatic binding energy observed in the optimizations. Side chains were built by performing a rotamer dihedral scan in CHARMM, using dihedral angle increments of 60 degrees, to determine the most desirable position for each side chain. Binding energies were then calculated for the wild type and mutant complexes using the Poisson-Boltzmann electrostatic energy and additional terms for the van der Waals energy and buried surface area.

The crystal structure of the α1 integrin I-domain (VLA-1) complexed with the Fab fragment of a humanized neutralizing antibody (AQC2) was solved to 2.8Å at a pH of 7.40. There were two complexes within the asymmetric unit cell. A manganese (MN) atom was at the complex interface in both complexes, with most of its interactions coming from the I-domain. Asp101 from the antibody mimics a collagen glutamate interaction.

The following table shows the optimization results obtained for CDR variable loop 2 in the heavy chain of AQC2. The Mut (Mutation energy) column corresponds to the binding free energy difference (in kcal/mol) in going from the native residue to a completely uncharged sidechain isostere, *i.e.*, a residue with the same shape but no charges or partial charges on the atoms. Negative numbers indicate a predicted increase of binding affinity. The Opt-1 column corresponds to the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of –1. The columns Opt0 and Opt1 correspond to the binding free energy

differences with optimal charges, the net charge being 0 and +1, respectively. Based on these results and the visual inspection of the structure, mutations are designed that can take advantage of these binding free energy improvements. For instance, the mutation from THR50 to VAL, which is an uncharged isostere, makes use of the predicted –0.52 kcal/mol in the mutation energy. The mutation LYS64 to GLU uses the –1.42 kcal/mol predicted maximal free energy gain for a mutation to a side chain with a net charge of –1.

The selection of mutant designs were further explored computationally according to the following rules.

For example, in those instances in which mutation energy (Mut, corresponding to the binding free energy difference (in kcal/mol) associated with a transition from the native residue to a completely uncharged side chain isostere, *i.e.*, a residue with the same shape but no charges or partial charges on the atoms) was modeled to be favorable (*e.g.*, $\Delta G < -0.25$ kcal/mol), mutations from the set of amino acids with nonpolar sidechains, *e.g.*, Ala, Cys, Ile, Leu, Met, Phe, Pro, Val were selected.

Where Opt-1 energy (corresponding to the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of –1) was favorable (*e.g.*, $\Delta G < -0.25$ kcal/mol), mutations from the set of amino acids with negatively charged side chains, *e.g.*, Asp, Glu were selected.

Similarly, where Opt+1 energy (corresponding to the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of +1) was favorable (*e.g.*, $\Delta G < -0.25$kcal/mol), mutations from the set of amino acids with positively charged sidechains, *e.g.*, Arg, His, Lys were selected.

Finally, in those cases in which Opt0 energy (corresponding to the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of 0) was favorable (e.g., $\Delta G < -0.25$kcal/mol), mutations from the set of amino acids with uncharged polar sidechains, *e.g.*, Asn, Cys, Gln, Gly, His, Met, Phe, Ser, Thr, Trp, Tyr, to which are added Cys, Gly, Met and Phe were selected.

Table 2 – Optimization results obtained for AQC2 CDR heavy chain variable loop 2

Number	Residue	Mut	Opt-1	Opt0	Opt1
50	THR	-0.52	0.3	-1.24	3.17
51	ILE	0	-1.05	-0.91	-0.56
52	SER	0.39	6.33	-0.09	1.77
53	GLY	---	---	---	---
54	GLY	---	---	---	---
55	GLY	---	---	---	---
56	HSD	-0.2	-0.09	-0.68	-0.02
57	THR	0.05	-0.77	-0.61	-0.3
58	TYR	-0.13	-1.98	-1.37	3.06
59	TYR	0.03	-1.35	-0.91	-0.39
60	LEU	0	-1.39	-1.08	-0.71
61	ASP	0.56	-0.11	0.25	0.64
62	SER	0.01	-0.21	-0.08	0.08
63	VAL	0	-0.98	-0.73	-0.36
64	LYS	-0.55	-1.42	-1.23	-0.97

As described before, the designed mutants are built *in silico* and the binding energy is recalculated. Results from these computational mutation calculations are shown below. Numbers represent change in binding affinity from wild-type to the mutant (negative meaning mutant is more favorable). Energies are the average of the two models.

Table 3 – Computational mutation calculations for AQC2 CDRs

Heavy Chain Modifications			
Mutation	Electrostatics	Full Energy	Type
Asp106Asn	-0.1	-0.1	3
Arg31Gln	-2.2	2.3	1
Arg31Glu	-0.8	4.9	1
Arg31Lys	0.5	2.7	1
Arg31Phe	0.9	2.8	2
Tyr32Phe	-0.4	0.6	2
Ser35Asn	-1.3	-1.3	2
Ser35Gln	-0.6	-0.7	2
Thr50Val	-1.2	-1.7	2
His56Phe	-0.8	-0.8	2
Tyr58Asn	-0.3	5.1	3
Tyr58Asp	-2.0	3.2	3
Tyr58Gln	-2.4	2.1	3
Tyr58Glu	-1.2	3.1	3
Tyr59Asp	-0.6	-0.6	3
Try59Glu	-0.5	-0.5	3
Leu60Asp	-0.1	-0.1	3
Leu60Glu	-0.3	-0.3	3
Lys64Asn	-0.6	-0.5	3
Lys64Asp	-0.9	-0.8	3
Lys64Gln	-0.6	-0.5	3
Lys64Glu	-0.9	-0.9	3

Table 3 – Computational mutation calculations for AQC2 CDRs (continued)

Light Chain Modifications			
Asn30Ala	-0.1	1.1	2
Asn30Ile	0.5	0.2	2
Asn30Leu	-0.3	-0.5	2
Asn30Val	-0.5	-0.2	2
His31Arg	1.6	1.9	1
His31Lys	-0.7	1.3	1
Leu49Arg	1.0	0.0	1
Leu49His	2.4	0.6	1
Leu49Lys	-0.1	-1.1	1
Asn52Arg	0.1	0.1	1
Asn52His	2.8	-0.2	1
Asn52Lys	0.3	1.5	1
Trp95Asp	2.5	4.4	3
Trp95Glu	0.7	2.9	3

As the results show, the computational process described above was successfully implemented to predict affinity enhancing side chain mutations. These findings were classified into three general classes of mutations. The first type of mutation involves residues at the interface across from a charged group on the antigen capable of making a hydrogen bond; the second involves buried polar residues that pay a desolvation penalty upon binding but do not make back electrostatic interactions; and the third involves long-range electrostatic interactions.

The first type of mutation is determined by inspection of basic physical/chemical considerations, as these residues essentially make hydrogen bonds with unsatisfied hydrogen partners of the antigen. Surprisingly, it was observed that the cost of desolvation seemed to outweigh the beneficial interaction energy in most cases. The second type of mutation represents a less intuitive type or set of mutations, as the energy gained is primarily a result of eliminating an unfavorable desolvation while maintaining non-polar interactions. The third mutation type concerns long-range interactions that show potential for significant gain in affinity. These types of mutations are particularly interesting because they do not make direct contacts with the antigen and should, therefore, pose less of a perturbation in the delicate interactions at the antibody-antigen interface.

In accordance with the computational data obtained as described above, mutants of ACQ2 (single chain Fv mutants) were generated, and their affinity was measured by the KinExA™ assay described above. The mutants generated to date are shown in the

table that follows. Where an affinity assay has been conducted, the results are shown in the column headed "Kd." The affinity of the original ACQ2 single chain Fv was 25 nM.

Table 4 – Observed affinity values for AQC2 altered antibodies

Heavy Chain Modifications	
Mutant	Kd
R31Q	8.2 nM
Y32F	34 nM
S35N	39 nM
S35Q	37 nM
T50V	14 nM
L60D	21 nM
K64E	38 nM
K64Q	12 nM
K64D	6.3 nM
K64N	4.1 nM
D106N	67 nM
Light Chain Modifications	
N30V	8.9 nM
H31R	31 nM
N52K	49 nM
N52R	17 nM
N52H	43 nM

The following alterations in AQC2 were also made: heavy chain modifications R31K, R31F, R31E, H56F, Y58E, Y58Q, Y59D, Y59E, L60E; light chain mutations N30L, N30A, N30I, H31K, L49K, L49R, L49H, W95E, W95D.

EXAMPLE 2
METHODS OF IMPROVING THE ANTIGEN-BINDING AFFINITY OF AN ANTI-CD154 ANTIBODY

In this example, methods for improving the binding affinity of an antibody against a therapeutically relevant antigen target, are described.

An antibody against human CD154 (also known as CD40 ligand or CD40L; see, *e.g.*, Yamada *et al.*, Transplantation, 73:S36-9 (2002); Schonbeck *et al.*, Cell. Mol. Life Sci. 58:4-43 (2001); Kirk *et al.*, Philos. Trans. R. Soc. Lond. B. Sci. 356:691-702 (2001); Fiumara *et al.*, Br. J. Haematol. 113:265-74 (2001); and Biancone *et al.*, Int. J. Mol. Med. 3(4):343-53 (1999)) which is a member of TNF family of proteins involved in mediating immunological responses, was raised by affinity maturation in mice. The 5c8 monoclonal antibody was developed from such studies and determined to inhibit the pathological processes mediated by CD154/CD40L.

In an effort to increase the affinity 5c8/CD40L interaction, electrostatic charge optimization techniques were applied to a crystal structure of the antibody-antigen complex in a two-level procedure to suggest improved-affinity mutants. First, electrostatic charge optimization was used to determine the position(s) of the CDR residue(s) that are sub-optimal for binding (Lee and Tidor, *J. Chem. Phys.* 106:8681-8690, 1997; Kangas and Tidor, *J. Chem. Phys.* 109:7522-7545, 1998). Second, a set of CDR mutations were determined for further computational analysis. Based on these calculations, the binding affinity was computationally determined for 23 modified antibodies having a single mutation (*i.e.*, 23 "single mutants"). It was predicted that 8 of the single mutants would be more favorable than wild-type antibody both in terms of electrostatic energy, and in terms of full energy function including a van der Waals energy term and a solvent accessible surface area term. These terms are unrelated to electrostatic forces, but they were calculated to ensure that the designed mutations did not contact other residues and would not reduce the amount of buried surface area significantly; increased buried surface area in complex formation is usually beneficial (see the "Full Energy" column of the table below).

The mutation predictions can be categorized as involving (1) mutations at the interaction interface involving residues that become partially buried upon binding (interactions are improved by making hydrogen bonds with the antibody); (2) mutations of polar residues on the antibody that become buried upon binding and thus pay a desolvation penalty but do not make any direct electrostatic interactions with the antibody (improvements are usually made by mutation to a hydrophobic residue with similar shape to the wild-type residue or by adding a residue that can make favorable electrostatic interactions); and (3) mutations of surface residues on the antibody that are in regions of uncomplementary potentials. These mutations are believed to improve long-range

electrostatic interactions between the antibody and antigen without perturbing packing interactions at the binding interface.

Based on results from a charge optimization, mutations were determined for computational analysis (the optimal charge distributions and design mutations that were closer to optimal than the current residue were examined; this process was done by inspection). A charge optimization gave charges at atom centers but did not yield actual mutation. A round of charge optimizations was performed with various constraints imposed to represent natural side chain characteristics. For example, an optimization was performed for a net side chain charge of –1, 0, and +1 with the additional constraint that no atom's charge exceeded an absolute value of 0.85 electron charge units.

The crystal structure of the CD40L/5c8 complex (PDB code: 1I9R) was prepared using standard procedures for adding hydrogens with the program CHARMM (Accelrys, Inc., San Diego, CA). N-acetamide and N-methylamide patches were applied to the N termini and C-termini, respectively. Using a continuum electrostatics model, an electrostatic charge optimization was performed on each side chain of the amino acids in the CDRs of the ACQ2 antibody. Appropriate side chain mutations were then determined based on the potential gain in electrostatic binding energy observed in the optimizations. Side chains were built by performing a rotamer dihedral scan in CHARMM, using dihedral angle increments of 60 degrees, to determine the most desirable position for each side chain. Binding energies were then calculated for the wild type and mutant complexes using the Poisson-Boltzmann electrostatic energy and additional terms for the van der Waals energy and buried surface area.

The crystal structure of the CD40 ligand complexed with the Fab fragment of a humanized neutralizing antibody (5c8) was solved to 3.1Å at a pH of 6.50. Since CD40L is naturally a trimer, there are three 5c8 Fab molecules and 5 CD40L molecules in the complex. They form three independent CD40L/5c8 interfaces in the complex. A zinc (ZN) atom was bound to each of the 5c8 Fab and it was included into the calculation. Calculations were carried out independently for three interfaces and the amino acid substitutions that were found to be favorable over wild type for all three sites were exploited.

The following table shows the optimization results obtained for CDR variable loop 1 in the light chain of 5c8 for all three 5c8 molecules. The Mut (Mutation energy) column corresponds to the binding free energy difference (in kcal/mol) in going from the native residue to a completely uncharged sidechain isostere, *i.e.*, a residue with the same shape but no charges or partial charges on the atoms. Negative numbers indicate a predicted increase of binding affinity. The Opt-1 column corresponds to the binding free energy difference that can be obtained with an optimal charge distribution in the side chain and a net side chain charge of –1. The columns Opt0 and Opt1 correspond to the binding free energy differences with optimal charges, the net charge being 0 and +1,

respectively. Based on these results and the visual inspection of the structure, mutations are designed that could take advantage of these binding free energy improvements. For instance, the mutation from SER 31 to VAL, which is an uncharged isostere, makes use of the predicted –1.23 to –0.98 kcal/mol in the mutation energy. The mutation GLN 27 to GLU uses the –1.21 to –0.88 kcal/mol predicted maximal free energy gain for a mutation to a side chain with a net charge of –1.

Table 5 – Optimization results obtained for 5c8 CDR light chain variable loop 1

Chain	Residue	Mut	Opt-1	Opt0	Opt1
1L	24 ARG	-0.11	0.17	-0.11	-0.37
1L	26 SER	-0.06	-0.59	-0.06	0.57
1L	27 GLN	0.21	-1.21	-0.95	-0.26
1L	28 ARG	0.11	-0.96	-0.71	-0.40
1L	30 SER	-0.01	-0.14	-0.42	-0.47
1L	31 SER	-1.23	3.88	-2.16	-0.42
1L	32 SER	1.45	0.91	-0.65	-0.67
1L	33 THR	-0.02	-0.66	-0.41	0.07
1L	34 TYR	-0.25	-1.00	-1.10	-0.80
1L	35 SER	-0.02	0.00	-0.11	0.04
1L	36 TYR	0.01	-0.95	-1.31	1.74
1L	38 HSD	-0.15	-0.48	-0.70	-0.62
2L	24 ARG	-0.46	-1.04	-0.46	0.13
2L	26 SER	-0.29	-1.60	-0.79	0.19
2L	27 GLN	0.26	-0.88	-0.41	0.35
2L	28 ARG	-0.59	-0.94	-0.46	0.08
2L	30 SER	0.08	-0.38	-0.55	-0.42
2L	31 SER	-0.98	4.04	-1.89	-0.54
2L	32 SER	0.74	2.31	-0.86	-0.87
2L	33 THR	0.00	-0.65	-0.38	0.09
2L	34 TYR	-0.09	-0.62	-0.48	-0.12
2L	35 SER	0.09	0.02	0.09	0.18
2L	36 TYR	0.10	-1.70	-1.24	2.37
2L	38 HSD	-0.23	-1.20	-1.17	-0.79
3L	24 ARG	-0.35	-0.34	-0.35	-0.35
3L	26 SER	-0.27	-1.23	-0.53	0.27
3L	27 GLN	0.11	-1.07	-0.71	-0.08
3L	28 ARG	-0.30	-0.85	-0.30	0.15
3L	30 SER	0.03	0.02	-0.29	-0.36
3L	31 SER	-1.06	4.02	-2.03	-0.90
3L	32 SER	0.82	1.18	-0.85	-1.05
3L	33 THR	0.20	-0.32	-0.15	0.29
3L	34 TYR	0.09	-0.80	-0.74	-0.38
3L	35 SER	0.06	-0.05	-0.10	-0.02
3L	36 TYR	0.04	-0.99	-1.30	1.66
3L	38 HSD	-0.20	-0.46	-0.76	-0.72

As described before, the designed mutants were built *in silico* and the binding energy was recalculated. Results from these computational mutation calculations are shown below. Numbers represent change in binding affinity from wild-type to the mutant (negative meaning mutant is more favorable). Energies for all three chains of 5c8 are given.

Table 6 – Computational mutation calculations for 5c8 CDRs

Chain Mutant	Full Energy	Electrostatics
1H TYR33PHE	0.197	-2.741
1H ASN59ASP	-0.995	-2.548
1H ASN59LEU	-1.294	-2.517
1L SER26ASP	-0.703	-0.712
1L GLN27GLU	-0.514	-0.357
1L SER31VAL	8.154	-1.739
1L THR33ASP	-0.219	-0.916
1L TYR54GLU	-0.999	-0.729
2H TYR33PHE	0.623	-2.726
2H ASN59ASP	-0.218	-2.885
2H ASN59LEU	-1.116	-3.067
2L SER26ASP	-1.333	-1.627
2L GLN27GLU	-0.658	-0.395
2L SER31VAL	9.293	-0.832
2L THR33ASP	-0.430	-1.359
2L TYR54GLU	-1.012	-1.030
3H TYR33PHE	0.145	-1.979
3H ASN59ASP	-0.837	-2.267
3H ASN59LEU	-1.179	-2.271
3L SER26ASP	-0.540	-0.565
3L GLN27GLU	-0.497	-0.342
3L SER31VAL	8.129	-1.284
3L THR33ASP	-0.337	-0.676
3L TYR54GLU	-1.123	-0.825

As the results show, the computational process described above was successfully implemented to predict affinity enhancing side chain mutations. These findings have been classified into three general classes of mutations. The first type of mutation involves residues at the interface across from a charged group on the antigen capable of making a hydrogen bond; the second involves buried polar residues that pay a desolvation penalty upon binding but do not make back electrostatic interactions; and the third involves long-range electrostatic interactions.

The first type of mutation was resolved by inspection, as these residues essentially make hydrogen bonds with unsatisfied hydrogen partners of the antigen. Surprisingly, the cost of desolvation seemed to outweigh the beneficial interaction energy in most cases. The second type of mutation represents a less intuitive type or set of mutations, as the energy gained is primarily a result of eliminating an unfavorable desolvation while maintaining non-polar interactions. The third mutation type concerns long-range interactions that show potential for significant gain in affinity. These types of mutations are particularly interesting because they do not make direct contacts with the antigen and, therefore, pose less of a perturbation in the delicate interactions at the antibody-antigen interface.

In accordance with the computational data obtained as described above, mutants of 5c8 (Fab fragments) were generated, and their affinity towards CD40L was measured by the KinExA™ assay described above. Selected results of some of the mutants generated to date are shown in the table that follows. Where an affinity assay has been conducted, the results are shown in the column headed "IC50." The affinity of the original 5c8 Fab to CD40L was 0.81 nM.

Table 7 – Observed affinity values for 5c8 altered antibodies

Mutant	IC50
Light	
S26D	0.26 nM
Q27E	0.12 nM

Accordingly, it was concluded that the methods of the invention allow for the affinity maturation of a an antibody of therapeutic relevance.

Equivalents

For one skilled in the art, using no more than routine experimentation, there are many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

(12) **United States Patent**
Tidor et al.

(10) **Patent No.: US 6,230,102 B1**
(45) **Date of Patent: May 8, 2001**

(54) **COMPUTER SYSTEM AND PROCESS FOR IDENTIFYING A CHARGE DISTRIBUTION WHICH MINIMIZES ELECTROSTATIC CONTRIBUTION TO BINDING AT BINDING BETWEEN A LIGAND AND A MOLECULE IN A SOLVENT AND USES THEREOF**

(75) Inventors: **Bruce Tidor**, Lexington; **Lee-Peng Lee**; **Sara E. Dempster**, both of Cambridge, all of MA (US)

(73) Assignee: **Massachusetts Institute of Technology**, Cambridge, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/055,475**

(22) Filed: **Apr. 3, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/042,692, filed on Apr. 4, 1997.

(51) **Int. Cl.**[7] G01N 33/48; G01N 33/50; G01N 33/00

(52) **U.S. Cl.** 702/19; 702/20; 436/89; 364/496; 364/578

(58) **Field of Search** 364/496, 578, 364/797, 499; 702/19, 20; 436/89

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,939,666	7/1990	Hardman	364/496
5,081,584	1/1992	Omichinski et al.	364/497
5,579,250	11/1996	Balaii et al.	364/496
5,612,895	3/1997	Balaii et al.	364/496

OTHER PUBLICATIONS

Brooks et al : Proteins., advances in chem. physics, vol. LXXVI, pp. 136–174, John Wiley, 1988.*

Gao, J. et al., "Hidden Thermodynamics of Mutant Proteins: A Molecular Dynamics Analysis", *Science*, vol. 244, pp. 1069–1072, Jun. 2, 1989.

Klapper, I. et al., "Focusing of Electric Fields in the Active Site of Cu–Zu Superoxide Dismutase: Effects of Ionic Strength and Amino–Acid Modification", Proteins: Structure Function and Genetics, 1986, pp. 47–59.

Wong, C.F. et al., "Cytochrome c: A Molecular Proving Ground for Computer Simulations", *J. Phys. Chem.*, vol. 97. No. 13, 1993, pp. 3100–3110.

DesJarlais, R. L., Robert P. Sheridan, J. Scott Dixon, Irwin D. Kuntz and R. Venkatarghavan, "Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape", *J. Med. Chem*, 289:2149–2153, 1986.

Connolly, Michael L., "Analytical Molecular Surface Calculation", *J. Appl. Cryst*, 16:548–558, 1983.

Richards, Frederic M., "Areas, Volumes, Packing, and Proteins Structure", *Ann. Rev. Biophys. Bioeng.*, 6:151–76, 1977.

Kuntz, Irwin D., Jeffrey M. Blaney, Staurt J. Oatley, Robert Langridge and Thomas E. Ferrin, "A geometric Approach to Macromolecule–Ligand Interactions", *J. Mol. Biol.*, 161:269–288, 1982.

Miranker, Andrew and Martin Karplus, "Functionality of Binding Sites: A Multiple Copy Simultaneous Search Method", *PROTEINS: Structure, Function, and Genetics*. 1:29–34, 1991.

Caflisch, Amedeo, Andrew Miranker And Martin Karplus, "Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV–1 Aspartic Proteinase", *J. Med. Chem.*, 36:2142–2167, 1993.

Eisen, Michael B., Don c. Wiley, Martin Karplus and Roderick E. Hubbard, "HOOK: A Program for Finding Novel Molecular Architectures That Satisfy the Chemical and Steric Requirements of a macromolecule Binding Site", *PROTEINS: Structure, Function, and Genetics*. 19:199–221, 1994.

Miranker, Andrew and Martin Karplus, "An Automated Method for Dynamic Ligand Design", *PROTEINS: Structure, Function, and Genetics*. 23:472–490, 1995.

Sitkoff, Doree, Kim A. Sharp and Barry Honig, "Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models", *J. Phys. Chem.*, 98:1978–1988, 1994.

Yang, An–Suei and Barry Honig, "Free Energy Determinants of Secondary Structure Formation: I. α–Helices", *J. Mol. Biol*, 252:351–365, 1995.

(List continued on next page.)

Primary Examiner—John S. Brusca
Assistant Examiner—Stephen Siu
(74) *Attorney, Agent, or Firm*—Wolf, Greenfield & Sacks, P.C.

(57) **ABSTRACT**

The present computer-implemented process involves a methodology for determining properties of ligands which in turn can be used for designing ligands for binding with protein or other molecular targets, for example, HIV targets. The methodology defines the electrostatic complement for a given target site and geometry. The electrostatic complement may be used with steric complement for the target site to discover ligands through explicit construction and through the design or bias of combinatorial libraries. The definition of an electrostatic complement, i.e., the optimal tradeoff between unfavorable desolvation energy and favorable interactions in the complex, has been discovered to be useful in ligand design. This methodology essentially inverts the design problem by defining the properties of the optimal ligand based on physical principles. These properties provide a clear and precise standard to which trial ligands may be compared and can be used as a template in the modification of existing ligands and the de novo construction of new ligands. The electrostatic complement for a given target site is defined by a charge distribution which minimizes the electrostatic contribution to binding at the binding sites on the molecule in a given solvent. One way to represent the charge distribution in a computer system is as a set of multipoles. By identifying molecules having point charges that match this optimum charge distribution, the determined charge distribution may be used to identify ligands, to design drugs, and to design combinatorial libraries.

6 Claims, 5 Drawing Sheets

OTHER PUBLICATIONS

Yang, An Suei and Barry Honig, "Free Energy Determinants of Secondary Structure Formation: II. β–Sheets", *J. Mol. Biol*, 252:366–376, 1995.

Friedman, Richard A. and Barry Honig, "A Free Energy Analysis of Nucleic Acid Base Stacking in Aqueous Solution", *Biophysical Journal*, 69:1528–1535, 1995.

Zhou, Zhongxiang, Philip Payne and Max Vasquez, "Finite–Difference Solution of the Poisson–Boltzmann Equation Complete Elimination of Self–Energy", *Journal of Computational Chemistry*, 11(11):1344–1351, 1996.

Klapper, Issac, Ray Hagstrom, Richard Fine, Kim A. Sharp and Barry H. Honig, "Focusing of Electric Fields in the Active Site of Cu–Zu Superoxide Dismutase: Effects of Ionic Strength and Amino–Acid Modification", *PROTEINS: Structure, Function, and Genetics*. 1:47–59, 1986.

Gilson, Michael K., Kim A. Sharp and Barry H. Honig, "Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment", *Journal of Computational Chemistry*, 9(4):327–335, 1987.

Luty, Brock A., Malcolm E. Davis and J. Andrew McCammon, "Solving the Finite–Difference Non–Linear Poisson–Boltzmann Equation", *Journal of Computational Chemistry*, 13(9):1114–1118, 1992.

Zacharias, Martin, Brock A. Luty, Malcolm E. Davis and J. Andrew McCammon, "Poisson–Boltzmann Analysis of the λ s Represor–operator Interaction", *Biophys J. Biophysical Society*, 63:1280–1258, Nov. 1992.

Zauhar, R. J. and R. S. Morgan, "The Rigorous Computation of the Molecular Electric Potential", *Journal of Computational Chemistry*, 9(2):171–187, 1988.

Bharadwaj, Ranganathan, Andreas Windemuth, S. Sridharan, Barry Honig and Anthony Nicholls, "The Fast Multipole Boundary Element Method for Molecular Electrostatics: An Optimal Approach for Large Systems", *Journal of Computational Chemistry*, 16(7):898–913, 1995.

* cited by examiner

FIG. 1

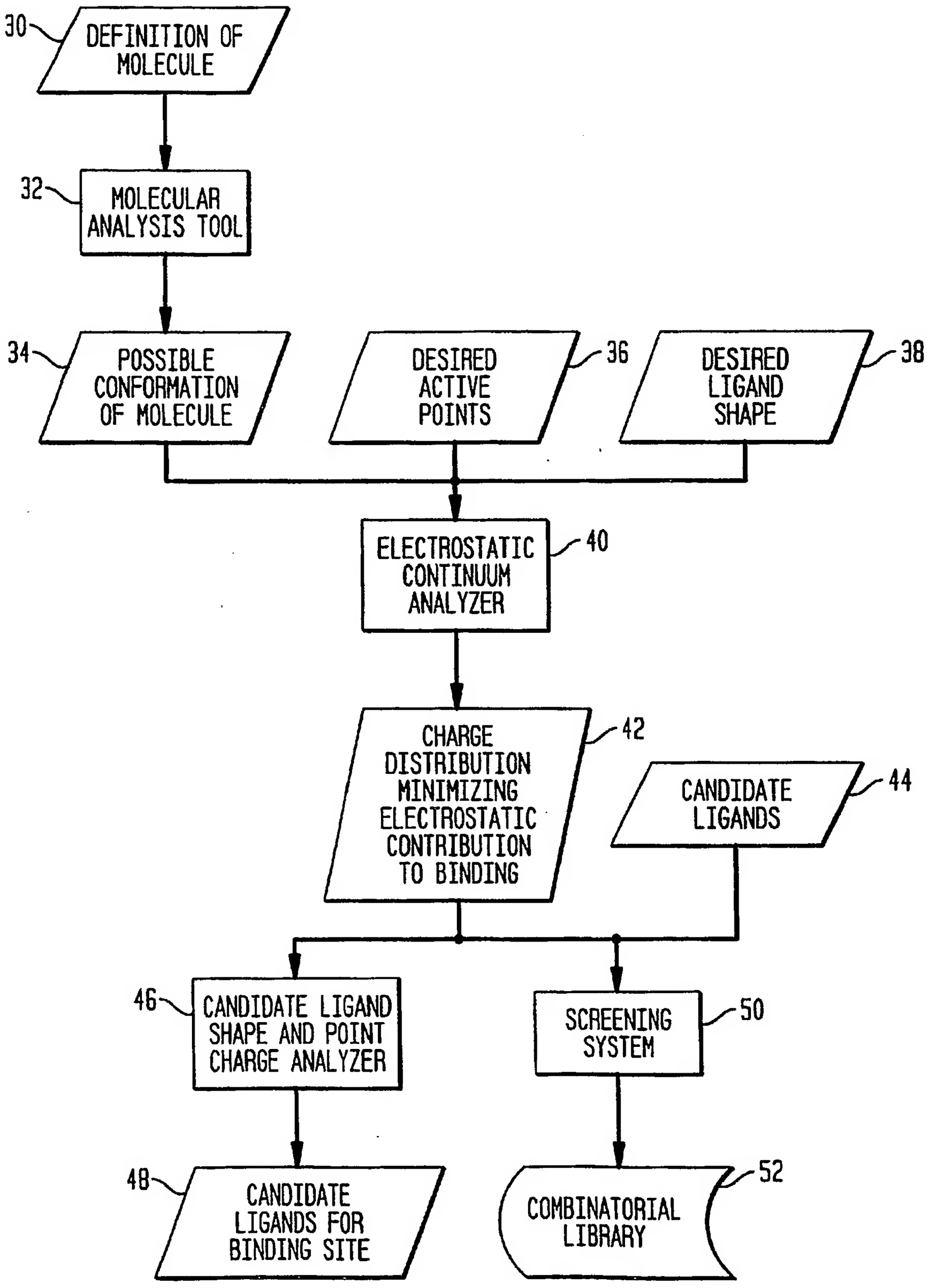

30
DEFINITION OF MOLECULE
32
MOLECULAR ANALYSIS TOOL
34
POSSIBLE CONFORMATION OF MOLECULE
36
DESIRED ACTIVE POINTS
38
DESIRED LIGAND SHAPE
40
ELECTROSTATIC CONTINUUM ANALYZER
42
CHARGE DISTRIBUTION MINIMIZING ELECTROSTATIC CONTRIBUTION TO BINDING
44
CANDIDATE LIGANDS
46
CANDIDATE LIGAND SHAPE AND POINT CHARGE ANALYZER
48
CANDIDATE LIGANDS FOR BINDING SITE
50
SCREENING SYSTEM
52
COMBINATORIAL LIBRARY

FIG. 2

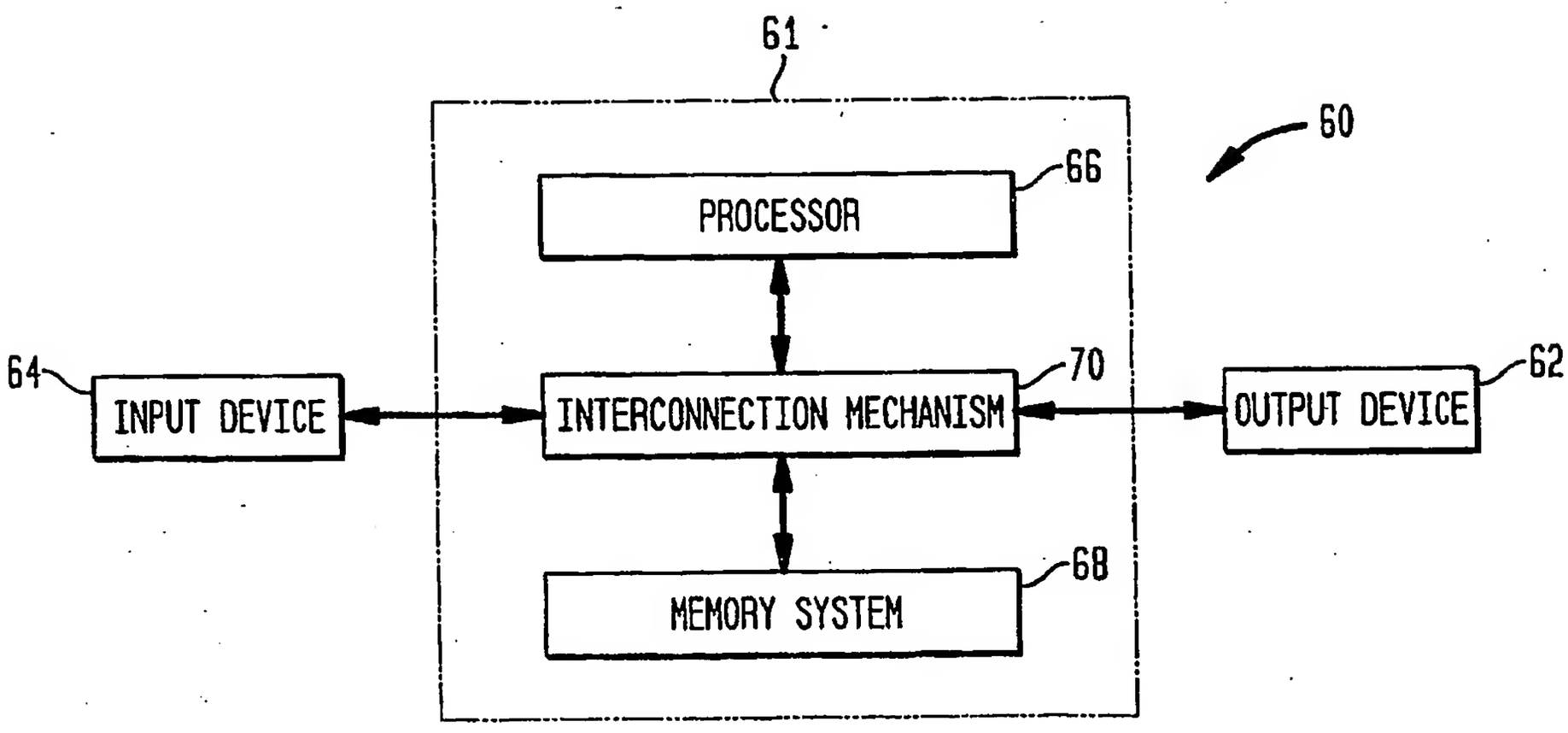

61
60
66
PROCESSOR
64
INPUT DEVICE
70
INTERCONNECTION MECHANISM
62
OUTPUT DEVICE
68
MEMORY SYSTEM

FIG. 3

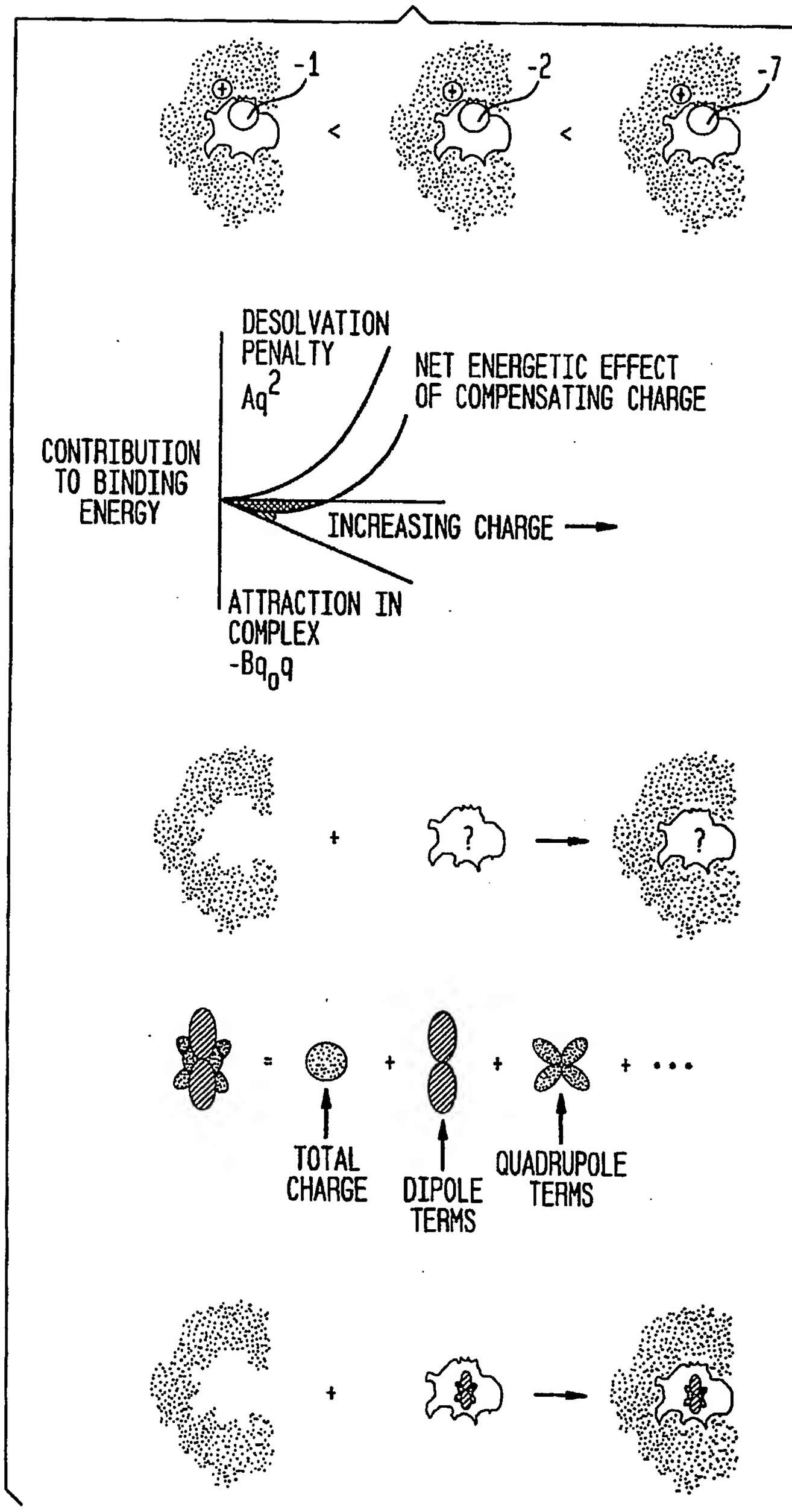
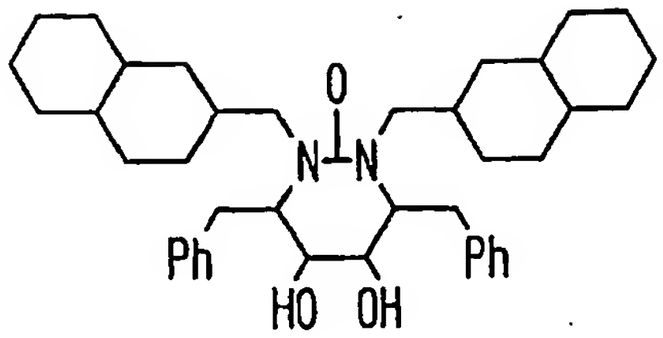

-1
<
-2
<
-7
DESOLVATION PENALTY
Aq^2
NET ENERGETIC EFFECT OF COMPENSATING CHARGE
CONTRIBUTION TO BINDING ENERGY
INCREASING CHARGE
ATTRACTION IN COMPLEX
$-Bq_0q$
?
?
=
+
+
+
+ • • •
TOTAL CHARGE
DIPOLE TERMS
QUADRUPOLE TERMS

FIG. 4A



a) XK 263
b) DMP 323
c) DMP 450

FIG. 4B

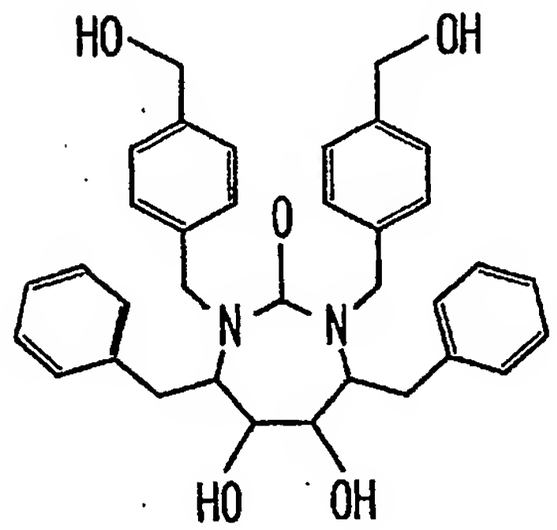


FIG. 4C

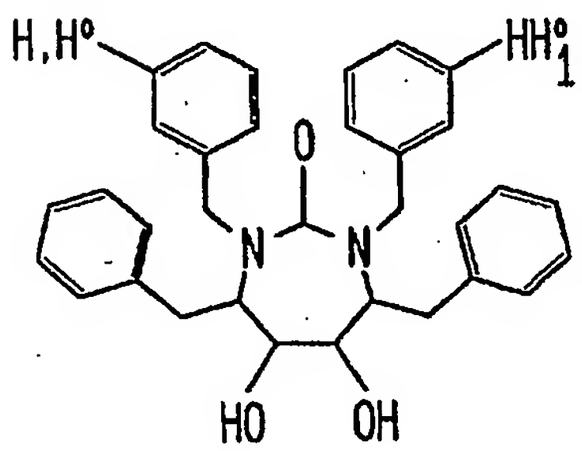
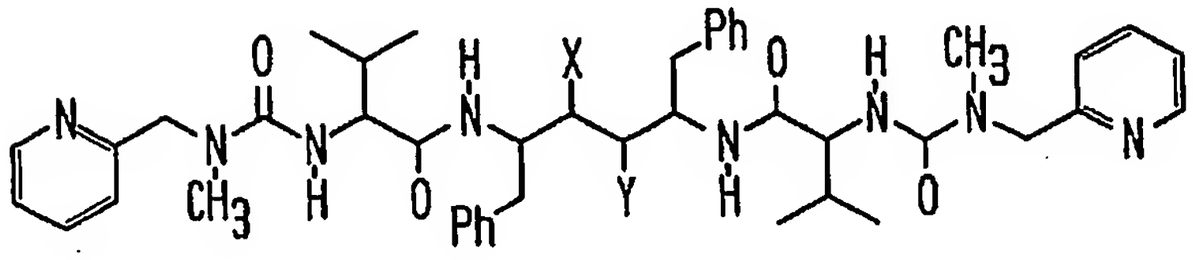


FIG. 4D



COMPOUND	X,Y
A-77003	R-OH, S-OH
A-76889	R-OH, R-OH
A-76982	S-OH, S-OH
A-78791	S-OH, H

FIG. 5A

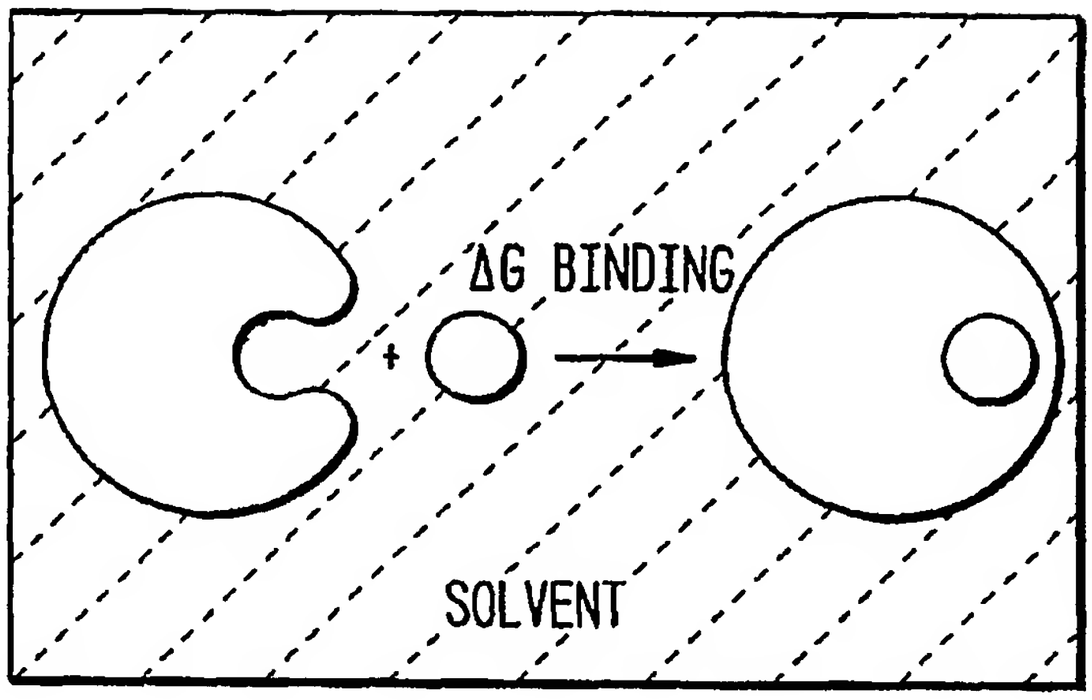
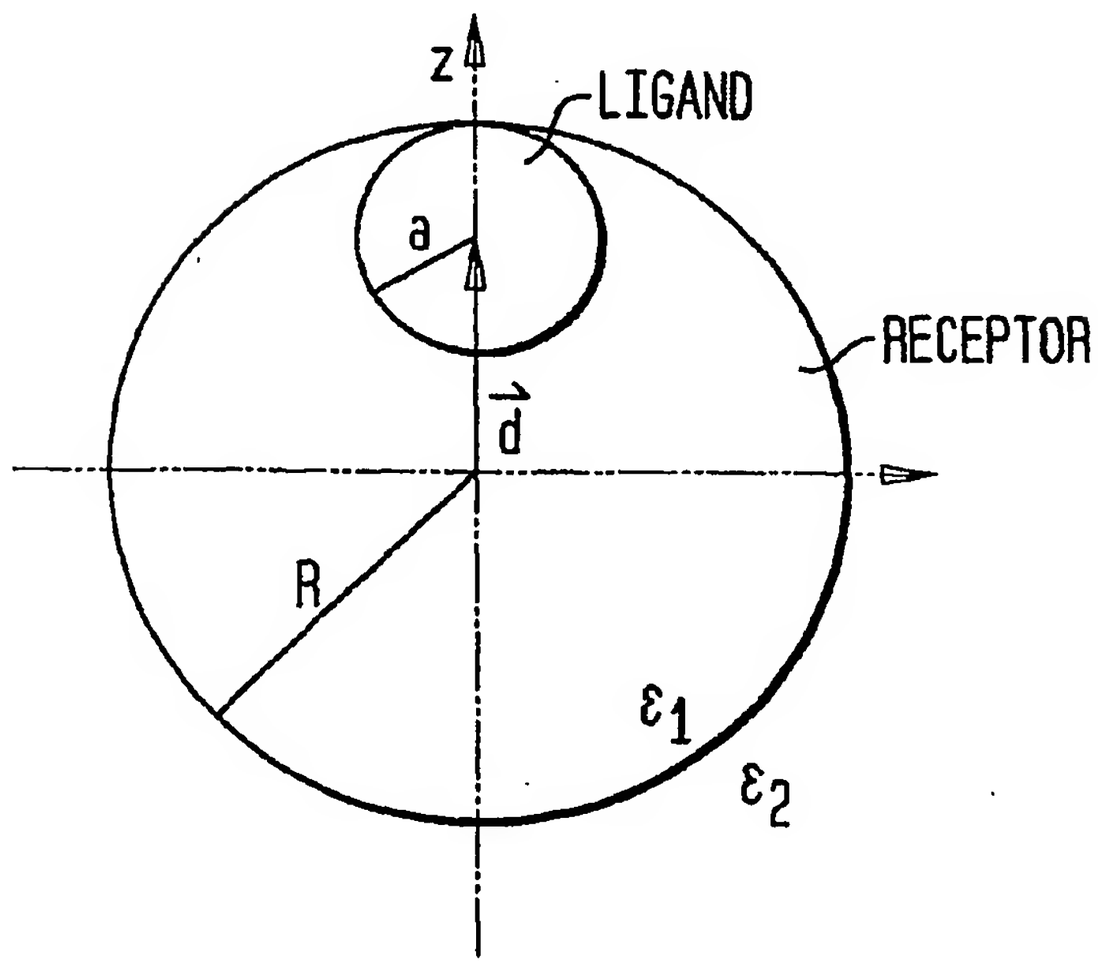

ΔG BINDING
SOLVENT

FIG. 5B


z
LIGAND
a
RECEPTOR
$\vec{d}$
R
ε_1
ε_2

COMPUTER SYSTEM AND PROCESS FOR IDENTIFYING A CHARGE DISTRIBUTION WHICH MINIMIZES ELECTROSTATIC CONTRIBUTION TO BINDING AT BINDING BETWEEN A LIGAND AND A MOLECULE IN A SOLVENT AND USES THEREOF

RELATED APPLICATIONS

This application claims the benefit under 35 USC §119(e) of U.S. Provisional Patent Application Serial No. 60/042,692 filed on Apr. 4, 1997, entitled COMPUTER SYSTEM AND PROCESS FOR IDENTIFYING A CHARGE DISTRIBUTION WHICH MINIMIZES ELECTROSTATIC CONTRIBUTION TO BINDING AT BINDING BETWEEN A LIGAND AND A MOLECULE IN A SOLVENT AND USES THEREFOR. The contents of the provisional application are hereby expressly incorporated by reference.

GOVERNMENT SUPPORT

This work was funded in part by grant numbers GM 47678 and GM 56552 from the National Institutes of Health. Accordingly, the United States Government may have certain rights to this invention.

FIELD OF THE INVENTION

The present invention relates to rational drug design, and more particularly, to rational drug design based upon the prediction of a charge distribution on a ligand which minimizes the electrostatic contribution to binding between the ligand and its target molecule in a solvent. The present process also relates to methods and tools for making such predictions and enhanced-binding ligands, and to the diagnostic and therapeutic uses of the ligands so produced.

BACKGROUND OF THE INVENTION

Methods for computational rational drug design include two general approaches: those that screen whole molecules and those that probe local sites and construct molecules through the joining of molecular fragments or grafting of chemical moieties onto a parent structure. DOCK is an example of a whole-molecule algorithm which uses a procedure to find the complementary shape to a given target site (I. D. Kuntz, et al., J. Mol. Biol. 161:269 (1982) (Kuntz); R. L. DesJarlais, et al., J. Med. Chem. 31:722 (1988) (DesJarlais)). Large compound databases can be computationally "screened" by first eliminating molecules whose shape is incompatible with the target site (by computing an overlap with the complementary shape) and then by attempting to rank those that remain with an approximate energy function. This procedure has been successful at identifying a number of ligands that bind to target sites. Unfortunately, X-ray crystal studies have shown that the ligands often bind differently in the site than predicted. One possible reason for this discrepancy between prediction and reality is that although the shape-complementarity algorithm is effective at removing extremely incompatible trial ligands, the approximate energy function is too inexact to define higher-level details of binding.

The MCSS (Multiple Copy Simultaneous Search) algorithm is one of the most popular fragment based approaches to ligand design (P. J. Goodford, J. Med. Chem. 28:849 (1985); A. Miranker and M. Karplus, Proteins: Struct., Funct., Genet. 11:29 (1991); and A. Caflisch, et al., J. Med. Chem. 36:2142 (1993)). The essential idea is to search the

region of a binding site and determine locations having especially favorable interaction energy with probes that represent a library of functional groups (carbonyl, amide, amine, carboxylate, hydroxyl, etc.). After the probes are successfully placed in the binding site, various subsets are linked to form coherent molecules. Two approaches to this problem have been developed. One attempts to fit small molecules from a database to join functional groups (HOOK) (M. B. Eisen, et al., Proteins: Struct., Funct., Genet. 19:199 (1995) and the other uses a simulated annealing protocol to grow linker atoms and bonds between fragments to produce ligand candidates with good covalent geometry and non-bonded interactions (DLD, dynamic ligand design) (A. Miranker and M. Karplus, Proteins: Struct., Funct., Genet. 11:29–34 (1991) and 23:472 (1995)).

The current methods for rational drug design are useful for suggesting novel and provocative geometries that appear to roughly compensate hydrogen-bonding groups in the site. Unfortunately, the current methods use approximations which may be inaccurate and which result in difficulties in accurately ranking candidates. Thus, although a number of computational algorithms exist both for the analysis of binding sites and bound complexes and for the rational design of ligands and other drug candidates, structure-based design remains an imprecise and non-deterministic endeavor.

SUMMARY OF THE INVENTION

The limitations of the prior art are overcome by providing for (i) a rigorous treatment of solvation, dielectric, and long-range electrostatic effects operating in both the unbound and the bound state of the target molecule and the ligand candidate, and (ii) a detailed quantitative method for ranking suggested ligands. The present process is based upon the discovery that the crude treatment of solvent, long-range electrostatics, and dielectric effects, as well as the lack of appropriate treatment for the unbound state of the target molecule and the ligand candidate, have limited the rational design and identification of novel ligand candidates for binding to a preselected target molecule. The present computer-implementation overcomes these limitations by providing a process which considers the exchange nature of ligand/target molecule binding, in which interactions with solvent are traded for interactions between a ligand and its complementary target molecule. In contrast to the prior art methods, the process disclosed herein takes into account solvent, long-range electrostatics, and dielectric effects in the binding between a ligand and its target receptor in a solvent.

Accordingly, in one aspect, a process for identifying properties of a ligand for binding to a target molecule (e.g., receptor, enzyme) in a solvent given a representation of a shape of the target molecule in three dimensions is provided. The process involves selecting a shape of the ligand defined in three dimensions, which shape is complementary to (matches) a shape of a selected portion of the target molecule; and determining a representation of a charge distribution on the ligand which minimizes the electrostatic contribution to binding between the ligand and the target molecule in the solvent. In some embodiments, the representation of the charge distribution is a set of multipoles. In other embodiments, the process further involves the step of identifying a molecule having point charges that match the representation of the charge distribution.

These methods are particularly useful for designing enhanced-binding ligands for binding to a target molecule

which has a known ligand. As used herein, an enhanced-binding ligand refers to a ligand which has a structure that is based upon that of a known ligand for the target molecule but which is modified in accordance with the methods disclosed herein to have a charge distribution which minimizes the electrostatic contribution to binding between the ligand and the target molecule in a solvent. Thus, the present computer-implemented process provides a method of rational drug design that identifies such improved ligands for binding to a target molecule having a known or predictable three-dimensional structure. The method involves selecting a shape of the ligand defined in three dimensions which matches a shape of a selected portion of the target molecule and determining a representation of a charge distribution on the ligand which minimizes electrostatic contribution to binding between the ligand and the target molecule in the solvent.

The target molecules for which ligands are identified using the claimed process are molecules for which a representation of the three dimensional shape of the molecule is known or can be predicted. Such target molecules include biopolymers and non-biopolymers. Exemplary biopolymers include proteins, nucleic acids, lipids, carbohydrates, and mixtures of the foregoing (e.g., glycoproteins, lipoproteins and so forth). Exemplary non-biopolymers include polyamides, polycarbonates, polyalkylenes, polyalkylene glycols, polyalkylene oxides, polyalkylene terphthalates, polyvinyl alcohols, polyvinyl ethers, polyvinyl esters, polyvinyl halides, polyvinylpyrrolidone, polyglycolides, polysiloxanes, polyurethanes, alkyl cellulose, polymers of acrylic and methacrylic esters, polyethylene, polypropylene, poly(ethylene glycol), poly(ethylene oxide), poly(ethylene terphthalate), poly(vinyl alcohols), polyvinyl acetate, polyvinyl chloride, polystyrene, polyvinylpyrrolidone, polymers of lactic acid and glycolic acid, polyanhydrides, poly(ortho) esters, polyurethanes, poly(butic acid), poly(valeric acid), poly(lactide-cocaprolactone) and copolymers thereof.

As used herein, the terms "protein" or "polypeptide" are used interchangeably to embrace a variety of biopolymers that are formed of amino acids, e.g., receptors, hormones, and enzymes. It should be understood that as described herein, references to a "protein", a "polypeptide", or a "receptor" are generally applicable to analogous structures, such as lipoproteins, glycoproteins, proteins which have other organic or inorganic groups attached, and multi-chain and multi-domain polypeptide structures such as large enzymes and viruses, and include non-biopolymers. In these instances, analogous issues regarding the electrostatic contribution to binding between the ligand and the protein molecule are present.

In some embodiments, the target molecule is a protein and the present computer-implemented process is used to identify novel and/or improved ligands for binding to a protein having a known three-dimensional structure in a solvent. Known binding partners of ligands and proteins include hormone/receptor, cofactor or inhibitor/enzyme, antigen/antibody, and so forth. For proteins to which a ligand previously has been identified, the present process is used to identify the appropriate modifications to the known ligand structure to achieve a charge distribution on the "improved" ligand that minimizes the electrostatic contribution to binding between the improved ligand and the protein compared to that of the unmodified (natural) ligand. Exemplary ligand/protein binding partners used as starting points for identifying "improved" ligands in accordance with the present process are provided in the examples.

In another aspect, a process identifying novel and/or enhanced-binding ligands that bind to a target molecule that

is a protein is provided. Proteins are known to fold into a three-dimensional structure which is dictated by the sequence of the amino acids (the primary structure of the protein) and by the solvent in which the protein is provided. The biological activity and stability of proteins are dependent upon the protein's three-dimensional structure. The three-dimensional structure of a protein can be determined or predicted in a number of ways. The best known way of determining a protein structure involves the use of X-ray crystallography. The three-dimensional structure of a protein also can be estimated using circular dichroism, light scattering, or by measuring the absorption and emission of radiant energy. Protein structure also may be determined through the use of techniques such as neutron diffraction, or by nuclear magnetic resonance (NMR). The foregoing methods are known to those of ordinary skill in the art and are described in standard chemistry textbooks (e.g., Physical Chemistry, 4th Ed. Moore, W. J., Prentiss-Hall N.J. (1972) and Physical Biochemistry, Van Holde, K. E., Prentiss-Hall, N.J. (1971)). Using the foregoing techniques, a number of recurring patterns in naturally occurring proteins have been identified, the most common of which are alpha helices, parallel beta sheets and anti-parallel beta sheets. See, e.g., R. Dickerson, et al., The Structure and Action of Proteins (1969). Together, the helices, sheets and turns of a protein's secondary structure produce the three dimensional structure of the active molecule. The three dimensional structure of proteins can be determined empirically using physical biochemical analysis or, alternatively, can be predicted using model building of three dimensional structures of one or more homologous proteins which have a known three dimensional structure.

The present computer-implemented process is particularly useful for designing an improved ligand that has a structure which is based upon the structure of a known ligand for a target molecule but which has been modified in accordance with the present methods to have a charge distribution which minimizes the electrostatic contribution to binding between the improved ligand and the target molecule in a solvent. Such improved ligands are referred to herein as "enhanced-binding ligands". Accordingly, the present process uses a ligand of known conformation as a starting point for the further optimization and selection of a ligand structure which will have reduced electrostatic contribution to binding to the molecule and the solvent. For example, the present process is used to produce an improved (enhanced-binding) co-factor or inhibitor of an enzyme (e.g., HIV-1 protease).

The present computer-implemented process also provides for the design of an improved hormone or other ligand for optimum binding (minimized electrostatic contribution to binding) to fit any known receptor site. This process is particularly useful for drug design, since it permits drugs to be designed and manufactured which more selectively and more stably are capable of binding to the receptor site. The design of improved ligands for binding to receptors means that lower dosages can be used, thereby reducing the chance of side effects and/or toxicity that may be associated with higher dosages. The design of improved ligands for binding to receptors also permits the identification of drugs having greater efficacy than the original ligand which is used as the basis for identifying an improved ligand having improved binding properties. Accordingly, known ligands for a protein can be used as a starting point for the design of improved ligands, wherein the improvement is based upon the improved binding properties of the ligand to the protein that are attributed to the selection of a ligand having a charge

distribution which minimizes the electrostatic contribution to binding between the ligand and the protein in a solvent. Thus, the present process permits the customizing of antigens and epitopes to more selectively and, with greater affinity, bind to antibodies, and also provides for the design and selection of novel and/or improved ligands which bind to other receptors or target molecules.

The ligands that are identified in accordance with the methods disclosed herein can be labeled with detectable labels such as radioactive labels, enzymes, chromophores and so forth for carrying out immuno-diagnostic procedures or other diagnostic procedures. These labeled agents can be used to detect the target molecules in a variety of diagnostic samples. For imaging procedures, in vitro or in vivo, the ligands identified herein can be labeled with additional agents, such as NMR contrasting agents or x-ray contrasting agents. Methods for attaching a detectable agent to a polypeptide or other small molecule containing reactive amino groups are know in the art. The ligands also can be attached to insoluble support for facilitating diagnostic assays.

The present computer-implemented process also is useful for searching three-dimensional databases for structures which have a shape which matches a shape of a selected portion of the protein and which also has a charge distribution which minimizes electrostatic contribution to binding between the ligand and the protein in a solvent. Alternatively, three-dimensional databases can be selected on the basis of the shape of the ligand alone (so that it matches a shape of a selected portion of the protein) with further modification of the database molecules that satisfy this criteria to have a charge distribution which minimizes electrostatic contribution to binding between the modified ligand and the protein in a solvent. Search algorithms for three-dimensional database comparison are available in the literature. See, e.g., U.S. Pat. No. 5,612,895, issued to V. Balaji, et al., "Method of Rational Drug Design Based on Ab Initio Computer Simulation of Conformational Features of Peptides" and references disclosed therein. For related computer methods for drug design, see also, U.S. Pat. No. 5,081,584, issued to Omichinski et al., "Computer-assisted Design of Anti-peptides Based on the Amino acid Sequence of a Target Peptide", and U.S. Pat. No. 4,939,666, issued to Hardman, "Incremental Macromolecule Construction Methods".

Each of the novel and/or "improved" ligands identified using the present process are prepared employing standard synthetic or recombinant procedures and then tested for bioactivity. Those compounds which display bioactivity are candidate peptidomimetics; those compounds which do not display bioactivity help further define portions of the ligand which are essential for binding of the ligand to the target molecule. As used herein, a peptidomimetics broadly refers to a compound which mimics a peptide. For example, morphine is a peptidomimetic of the peptide endorphin.

A database of known compounds (e.g., the Cambridge Crystal Structure Data Base, Crystallographic Data Center, Lensfield Road, Cambridge, CB2 1EW, England; and Allen, F. H., et al., Acta Crystallogr., B35:2331 (1979)) also can be searched for structures which contain the steric (shape) parameters used for complementing (matching) a shape of a selected portion of the target molecule. Compounds which are found to contain the desired steric parameters are retrieved, and further analyzed to determine which of the retrieved compounds also have the desired charge distribution or that can be modified to have the desired charge distribution to minimize the electrostatic contribution to

binding between the ligand and the target molecule in a solvent. Ligands that are found to contain both the desired shape and charge distribution are additional candidates as peptidomimetics of the original target peptides.

The ligands which are identified in accordance with the present computer-implemented process are evaluated for biological activity and/or for binding affinity to the target molecule. An iterative approach is used to identify the ligands having the most preferred biological properties. See, e.g., PCT WO 19359, "Process for making Xanthene or Cubane based compounds, and Protease Inhibitors", which describes an iterative process for identifying the bioactive conformation of an enzyme inhibitor in a complex chemical combinatorial library. The bioactive conformation then is used to design peptidomimetics, or used to search a three-dimensional database of organic structures to suggest potential peptidomimetics. Standard physiological, pharmacological and biochemical procedures are available for testing the "improved" or novel ligands identified using the present process. The particular protocol for evaluating bioactivity is a function of the compound that is being tested. This kind of analysis can be applied to known ligands that bind to a target molecule, (e.g., HIV protease, MHC class II proteins) to design enhanced-binding ligands for these biologically important target molecules.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings,

FIG. 1 is a block diagram describing one embodiment of the present computer-implemented process;

FIG. 2 is a block diagram of a computer system which may be used to implement the present computer-implemented process;

FIG. 3 is a diagram illustrating chemical principles underlying the present computer-implemented process; and

FIG. 4 is a diagram illustrating inhibitors of HIV-1 protease.

FIG. 5 is an illustration of problem geometries.

DETAILED DESCRIPTION

The present computer-implemented process will be more completely understood through the following detailed description which should be read in conjunction with the attached drawing in which similar reference numbers indicate similar structures. All references cited above and in the following description are hereby expressly incorporated by reference.

The the present computer-implemented process involves a methodology for determining properties of ligands which in turn is used for designing ligands for binding with protein or other molecular targets, for example, HIV targets. The methodology defines the electrostatic complement for a given target site and geometry. The electrostatic complement may be used with a steric complement for the target site to discover ligands through explicit construction and through the design or bias of combinatorial libraries.

The definition of an electrostatic complement, i.e., the optimal tradeoff between unfavorable desolvation energy and favorable interactions in the complex, has been discovered to be useful in ligand design. This methodology essentially inverts the design problem by defining the properties of the optimal ligand based on physical principles. These properties provide a clear and precise standard to which trial ligands may be compared and used as a template in the modification of existing ligands and the de novo construction of new ligands.

The electrostatic complement for a given target site is defined by a charge distribution which minimizes the electrostatic contribution to binding at the binding sites on the molecule in a given solvent. One way to represent the charge distribution in a computer system is as a set of multipoles. By identifying molecules having point charges that match this optimum charge distribution, the determined charge distribution may be used to identify ligands, to design drugs, and to design combinatorial libraries.

Referring now to FIG. **1**, one embodiment of the present computer-implemented process is shown. This embodiment may be implemented using one or more computer programs on a computer system, an example of which is described below. Given a definition of a molecule for which a ligand is to be designed, indicated at **30**, a molecular analysis tool **32** provides a possible conformation, or shape, of the molecule as indicated at **34**. There are several systems available to provide such conformations, including but not limited to x-ray crystallography, homology modeling, nuclear magnetic resonance imaging or analytical techniques such as shown in Kuntz and DesJarlais. The desired binding or active points on the molecule, indicated at **36**, and a desired ligand shape for binding with the molecule at the indicated binding points, as indicated at **38**, also are input to the computer system.

An electrostatic continuum analyzer **40**, described in more detail below, is used to determine a charge distribution which minimizes the electrostatic contribution to binding at the binding sites in a given solvent, given the representation of the shape of the molecule in three dimensions, the binding sites on the molecule defined by locations in three dimensions and the desired ligand shape, also defined in three dimensions. Accordingly, the output of analyzer **40** is a representation of a charge distribution minimizing electrostatic contribution to binding as indicated at **42**.

The charge distribution **42** is used in combination with candidate ligands having the desired ligand shape, as indicated at **44**. A candidate ligand shape and point charge analyzer **46** determines which candidate ligands have a charge distribution closest to the optimal charge distribution **42**. Analyzer **46** outputs candidate ligands for the binding site as indicated at **48**. Similarly, a screening system **50** may also be used to screen candidate ligands **44** for their proximity to the optimum charge distribution indicated at **42** in order to develop a combinatorial library **52**. Such a combinatorial library may be used to develop more complex molecules having desired characteristics.

Referring now to FIG. **2**, a suitable computer system **60** typically includes an output device **62** which displays information to a user. The computer system includes a main unit **61** connected to the output device **62** and an input device **64**, such as a keyboard. The main unit **61** generally includes a processor **66** connected to a memory system **68** via an interconnection mechanism **70**. The input device **64** is also connected to the processor **66** and memory system **68** via the connection mechanism **70**, as is the output device **62**.

It should be understood that one or more output devices may be connected to the computer system. Example output devices include a cathode ray tube (CRT) display, liquid crystal displays (LCD), printers, communication devices such as a modem, and audio output. It should also be understood that one or more input devices may be connected to the computer system. Example input devices include a keyboard, keypad, track ball, mouse, pen and tablet, communication device, audio input and scanner. It should be understood present computer-implemented process is not

limited to the particular input or output devices used in combination with the computer system or to those described herein.

The computer system **60** may be a general purpose computer system which is programmable using a high level computer programming language, such as "C", "Fortran," or "Pascal." The computer system may also be specially programmed, special purpose hardware. In a general purpose computer system, the processor is typically a commercially available processor, of which the series x86 processors, available from Intel, and the 680X0 series microprocessors available from Motorola are examples. Many other processors are available. Such a microprocessor executes a program called an operating system, of which UNIX, DOS and VMS are examples, which controls the execution of other computer programs and provides scheduling, debugging, input/output control, accounting, compilation, storage assignment, data management and memory management, and communication control and related services. One embodiment was implemented using a Hewlett-Packard 9000/735 computer with a PA-7200 (99 MHz) chip. The processor and operating system define a computer platform for which application programs in high-level programming languages are written.

A memory system typically includes a computer readable and writeable nonvolatile recording medium, of which a magnetic disk, a flash memory and tape are examples. The disk may be removable, known as a floppy disk, or permanent, known as a hard drive. A disk has a number of tracks in which signals are stored, typically in binary form, i.e., a form interpreted as a sequence of one and zeros. Such signals may define an application program to be executed by the microprocessor, or information stored on the disk to be processed by the application program. Typically, in operation, the processor causes data to be read from the nonvolatile recording medium into an integrated circuit memory element, which is typically a volatile, random access memory such as a dynamic random access memory (DRAM) or static memory (SRAM). The integrated circuit memory element allows for faster access to the information by the processor than does the disk. The processor generally manipulates the data within the integrated circuit memory and then copies the data to the disk when processing is completed. A variety of mechanisms are known for managing data movement between the disk and the integrated circuit memory element, and the present process is not limited thereto. It should also be understood that the present process is also not limited to a particular memory system.

It should be understood the present computer-implemented process is not limited to a particular computer platform, particular processor, or particular high-level programming language. Additionally, the computer system **60** may be a multiprocessor computer system or may include multiple computers connected over a computer network.

Defining Ligand Properties

The process for defining complementary ligand properties of electrostatic interactions, using such continuum calculations is outlined in FIG. **3**. Because of the exchange nature of electrostatic interactions, seemingly "strong" electrostatic attractions found in the bound state frequently destabilize the binding equilibrium, but presumably contribute to specificity. That is, because of the substantial desolvation penalty incurred for burying polar and charged groups, their net electrostatic contribution to macromolecular association is generally unfavorable. In designing a ligand for a given, fixed target that has polar and charged groups at the site, it is important to balance the desolvation and interaction

energies so as to contribute most favorably to binding or at least to provide the smallest possible destabilization. The following method solves this problem analytically, using idealized geometries and a continuum electrostatic model, and provides a single, unique optimum which is solved exactly.

For the case of binding a spherical ligand to an arbitrarily shaped receptor to form a spherical complex, the free energy of binding is expressed in terms of the charge multipoles of the ligand. By minimizing the binding free energy with respect to the multipoles, (i) there is a single, optimal multipole distribution defining the tightest binding ligand for the given geometry, (ii) this multipole distribution corresponds to a minimum in $\Delta G_{binding}$, (iii) at this optimum the magnitude of the ligand desolvation penalty is exactly half the magnitude of the favorable intermolecular electrostatic interactions in the complex, and (iv) the loss in binding free energy for a sub-optimal charge distribution is easily calculated by comparing to the optimum. This minimization of the binding free energy with respect to the multipoles provides a clear and unambiguous route from the continuum model, an accepted energetic description of macromolecules and ligands to a set of descriptors, i.e., the multipoles, for the optimal ligand. For this method to be broadly applicable, any requirement for spherical geometries is removed. Accordingly, macromolecules and ligands are of arbitrary shape and are treated as such.

In the spherical case, a variational binding energy for optimization, is defined as follows:

$$\Delta G_{var}=\Delta G_{int,L-R}+G_{hyd,L}{}^{bound}-G_{hyd,L}{}^{unbound} \quad (1)$$

This includes three terms, which are discussed separately here. The first is the ligand-receptor screened interaction energy, which includes a contribution from the interaction of each multipole component of the ligand charge distribution with all point charges in the receptor. These contributions are accounted for by coefficients, the $\alpha_{l,m}$, which are computed analytically, and the ligand multipoles ($Q'_{l,m}$),

$$\Delta G_{int,L-R}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q'^{*}_{l,m} \quad (2)$$

The second term is the bound-state reaction-field energy due to the ligand charge distribution, $G^{bound}_{hyd,L}$. It has contributions from all pairs of multipole components with the same value of m, since the ligand multipole distribution is generally expanded about a point that is not the center of the spherical boundary in the bound state but the geometry is chosen with azimuthal symmetry,

$$G^{bound}_{hyd,L}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\beta_{l,m,l',m}Q'^{*}_{l,m}Q'_{l',m} \quad (3)$$

The third term is the unbound-state solvation energy, which involves a contribution from each multipole component. Because the multipole expansion is taken about the center of the ligand sphere, and due to the orthogonality of the spherical harmonics, all cross-terms cancel, giving

$$G^{unbound}_{hyd,L}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q'^{*}_{l,m}Q'_{l,m} \quad (4)$$

Combining the above three equations,

$$\Delta G_{var}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q'^{*}_{l,m}+\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}(\beta_{l,m,l',m}-\delta_{l,l'}\gamma_{l,m})Q'^{*}_{l,m}Q'_{l',m} \quad (5)$$

and transforming to matrix notation, one completes the square and solves for the $Q'^{opt}_{l,m}$ giving the optimal variation binding energy. Since terms neglected from the variational binding energy are constant for a given geometry, these describe the multipoles of the optimal binding ligand. A more detailed exposition of this process is set forth in the article by L. P. Lee and B. Tidor, J. Chem. Phys., 106:8681–8690, (1997), which is expressly incorporated herein by reference and part of which can also be found in the Appendix.

An implementation of a computer program to perform the kinds of processing outlined in Lee and Tidor, supra and in the Appendix, can receive as an input the value l_{max}, which determines the size of the matrix of equations 59 and 61 (see Appendix), l_{cur}, which truncates the innermost summation in equations 25 and 46 (see Appendix), the geometry of the problem, which indicates the shape of the target and the ligand, and whether the monopole of the optimum is to be free or fixed at some value. The geometry of the problem includes the radius and coordinates of the center of both the bound state and ligand spheres on the z axis and the coordinates of magnitude of each partial atomic charge in the system. The dielectric constants $\in_1$ and $\in_2$ are determined by the particular problem. Evaluation of the α_i, β_{ij} and γ_i values is carried out, followed by solution of matrix equation 59 or 61 (see Appendix), for example by using LU decomposition. The eigen values of the B matrix may be obtained to verify that the stationary point was a minimum. All real floating point values may be represented, for example, using 64 bits or other suitable format. The computation of the matrix algebra may be accomplished using available or increased precision versions of appropriate subroutines, such as defined in Press et al, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1988. The output of the program when executed is a representation of the optimal charge distribution (e.g., using multipoles), the nature of the stationary point and a file recording the alpha, beta and gamma values. Because a direct method, i.e., LU decomposition, was used to solve a matrix equation, the time scales as $(l_{max})^6$ and the memory used scales as $(l_{max})^4$. This program output may be improved by accounting for the particularly sparse matrix in the matrix equation. Optimizations also may be provided by solving the matrix equation with iterative methods such as the conjugate gradient method or various relaxation methods. This method has been implemented and tested using both a highly symmetric charge distribution and the terminus of an alpha-helix as the receptor.

This method is extended to arbitrarily shaped molecules, by using iterative numerical computation to calculate the corresponding matrix coefficients and, for efficiency, by using a number of centers dispersed through the ligand volume at which individual multipole expansions are located.

When this method is extended to non-spherical geometries, it takes the following form. The $\alpha_{l,m}$ retain the same character, the $\beta_{l,m,l',m}$ become $\beta_{l,m,l',m'}$ because azimuthal alignment can no longer be used and the $\gamma_{l,m}$ become

$\gamma_{l,m,l',m'}$ because the orthogonality of the spherical harmonics does not eliminate the cross-terms for a non-spherical surface. Thus, a very similar matrix equation is found,

$$\Delta G_{var} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q_{l,m}^{*}+\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}(\beta_{l,m,l',m'}-\gamma_{l,m,l',m'})Q_{l,m}^{*}Q_{l',m'} \quad (6)$$

which is solvable by the same matrix methods used for the spherical case or by singular value decomposition using available or improved precision versions of appropriate subroutines, such as defined in Press et al, supra. However, numerical computations may be used to calculate the corresponding matrix coefficients. For the spherical case above, closed form expressions may be derived for rapid computation. When the same matrix coefficients ($\alpha_{l,m}$, $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$) are computed using iterative numerical methods, the computing requirements increase substantially.

Alternatively, the ligand may be described by using more centers, each described by a small number of multipoles. In the extreme, each ligand can be composed on point-charge locations, and currently 500 would be affordable, i.e., computable in under three weeks of computer time, with l_{max}=0 at each center. It is likely that the best solution will be intermediate, in which there are roughly 10 locations with l_{max}=2 (monopole, dipolar, and quadrupolar terms) or so at each center. The distributed centers of low-order multipoles may be an efficient and accurate way to describe arbitrary ligand charge distributions. The method has been properly elaborated for inclusion of interactions between separate multipole centers, and results using spherical geometries indicate that using two multipole distributions rather than one allows an equivalent description of the optimal charge distribution to be achieved using roughly one-quarter the number of multipole components and thus essentially one-quarter the time.

Two alternative schemes may be used for iterative numerical computation of matrix coefficients. The first scheme is a modification of a finite-difference Poisson-Boltzmann (FDPB) solver, such as DELPHI (I. Klapper, R. Hagstrom, R. Fine, K. Sharp. and B. Honig). Focusing of electric fields in the active site of Cu—Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. Proteins: Struct., Funct., Genet. 1: 47–59 (1986), M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. J. Comput. Chem. 9: 327–335 (1988) and UHBD (B. A.

Luty, M. E. Davis, and J. A. McCammon. Solving the finite-difference non-linear Poisson-Boltzmann equation. J. Comput. Chem. 13: 1114–1118 (1992), M. Zacharias, B. A. Luty, M. E. Davis, and J. A. Mcammon. Poisson-Boltzmann analysis of the λ repressor-operating interaction. Biophys. J. 63: 1280–1285 (1992)), and the second scheme is a modification of boundary-element methods (BEM) (R. J. Zauhar and R. S. Morgan. The rigorous computation of the molecular electric potential. J. Comput. Chem. 9: 171–187 (1988), R. Bharadwaj, A. Windemuth, S. Sridharan, B. Honig, and A. Nicholls. The fast multipole boundary element method for molecular electrostatics: An optimal approach for large systems. J. Comput. Chem. 16: 898–913 (1995)). These modifications allow point multipoles, as opposed to just point charges, to be represented.

Thus, a more complex method includes the following. For each pole component at each center iterative continuum calculations are carried out to determine its screened coulombic interaction with the receptor point charges (essentially $\alpha_{l,m}$), its interaction with its own reaction field and that of each of the other pole components in the system in the bound (essentially $\beta_{l,m,l',m'}$) and unbound (essentially $\gamma_{l,m,l',m'}$) state. It is estimated that a ligand represented by 99 pole components (such as 11 centers with l_{max}=2 at each) will require under three weeks of CPU time. For many applications, half that number of centers and half the time could be sufficient. A multipole distribution about a single center uses many global terms to accurately describe a complex charge distribution fairly far from the center. By distributing in space a number of centers for the expansion, an equally accurate description can be obtained with many fewer, somewhat more local, terms.

Using Molecular Descriptors to Discover Ligands

Referring again to FIG. 1, the charge distribution 42 defined by the above procedure may be used to determine which candidate ligands would have a charge distribution closest to the optimum. The descriptions of the charge distribution and molecular shape can be used to construct ligand structures de novo, or they can be used to screen compound databases, or they can be used in the design or bias of combinatorial libraries.

In the process of discovering ligands, detailed point-charge distributions are fit to the multipole distributions determined using the above methods. Next, molecular fragments and/or molecules are fit to the point-charge distributions. Finally, both the point charges and the fragments may be used in the design of combinatorial libraries, described below.

A least squares fitting procedures may be used to define a point-charge distribution that is a close fit to the multipole distribution describing the optimum. For example, a regular cubic lattice of grid points with roughly the spacing used in FDPB computations may be used. This can be achieved using the same tri-linear function used in FDPB codes to carry out the mapping in the opposite direction arbitrary point charges mapped to charged lattice points. (See Klapper.) Whether a set of point charges can provide an adequate fit to the electrostatic charge distribution represented by the multipoles, can be determined by comparing the decrease in free energy of binding due to using the fit point-charge distribution in place of the multipoles themselves. A trial using a cubic lattice of grid points with 0.5-Å spacing indicates that the computed loss in binding energy is less than 0.001 kcal/mol due to fitting point charges. In addition, the resulting point charges assigned are reasonable in magnitude (nearly all are less than 0.10e), making a fit to molecular fragments plausible. In this embodiment the multipoles, which are a somewhat non-local description of the charge distribution, are converted into a local grid based point-charge description so molecules can be fit.

The effectiveness of set points in fitting charges may be measured not only by minimizing the loss in binding energy, but also by how simply molecules or molecular fragments may be constructed from the point charges. The cubic lattice is used as described above to fit functional groups and molecules. A more molecule-based grid may also be used and may include connectivity for the common valencies (sp sp^2, and sp^3) co-embedded. Additionally, a uniform density of point charges may be a disadvantage, rather, having a higher density of point charges near the ligand surface may provide a more effective fit.

Given a point-charge distribution, it may be fit to a molecule or molecular fragment in several ways. For example, a database of molecular fragment geometries and point-charge distributions (such as a library derived from the

PARSE parameter set of fragments (D. Sitkoff, K. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. 98: 1978–1988 (1994)), may be used to match individual functional groups to favorable locations on the point-charge distribution. This matching process could be a very large scanning search if each fragment needed to be attempted at all locations and in every orientation in the ligand volume. The timing may be improved substantially, though, using a regular cubic lattice for the point-charge distribution. Each fragment would only need to be scanned over a relatively small section of lattice to determine sets of lattice point charges "diagnostic" for it. These diagnoses may be compiled for all library fragments, for example, in a hash table, and clusters of point-charge values may be used to query the hash table and fit fragments to the charge lattice. So long as the same grid spacing is maintained, the hash table may be reused for many different targets and optimizations.

After fragments have been placed, the problem of fitting them together into molecules is similar to the one addressed by the MCSS algorithm described above, although the theoretical foundations for choosing fragment locations are very different in that method and in the present computer-implemented process. Two solutions developed there may be adapted for use here. In the HOOK method described above, a database of small molecules is used as linkers to fit fragments together, generally trying to introduce rigidity at the same time. In the dynamic ligand method (DLD) described above, a sea of carbon atoms is superposed with the fragments and a simulated annealing procedure is used in which the occupancy of each carbon can grow and shrink and in which bond-making and bond-breaking events are used to coalesce novel carbon linkers. In each method, each fragment generally is allowed to move somewhat in order to create relatively unstrained ligands. An accurate penalty function for movement based on how movement affects the computed binding energy may be used. The DLD based approach may be better because of its flexibility.

Alternatively, molecules may be grown in a sequential fashion so as to fill the ligand volume and fit the point-charge distribution. A straightforward scheme involves placing a single fragment at a location where it fits the point-charge field and executing a search for other fragments that can be joined to the first, adjusting their relative orientation via the connecting torsion. This procedure can be carried out in a tree-like manner to create large numbers of ligands. An appropriate figure of merit or distance metric, is applied to determine whether to accept or reject each new fragment. A potential that includes van der Waals and torsional terms as well as a fit to the volume and charge distribution of the defined optimum may be used in this method.

Yet another alternative is the design of "minimalist" ligands. The multipole distributions of the optima may be fit with as few point charges as possible. This optimization process involves finding a relatively small number of point charges whose computed binding energy is within a few tenths of a kcal/mol of the optimum. Studies with complementary nucleotide bases suggest that a better complementary "base" than that used by nature can be reconstructed using only one-third the number of pont charges, i.e., a complement to adenine can be constructed using only four point charges; and this complement binds tighter than adenine. These reduced point-charge ligands retain the Watson-Crick hydrogen bonding to the partner, although in somewhat different fashion. Models for ligands containing very few required point charges may represent the key compensating interactions that a more elaborate ligand

should contain. They may serve as useful scaffolds or seeds upon which further computational molecular design should be carried out or about which a synthetic combinatorial strategy could usefully be built. If they bind tightly enough, they may be particularly useful therapeutics because it may be very difficult for the virus to evolve resistance to a small ligand targeted to catalytic side chains.

Designing Combinatorial Libraries

The design of combinatorial libraries as illustrated at **50** in FIG. 1, will now be described in more detail. Although there is substantial long-standing interest in using computational molecular modeling to carry out de novo rational ligand design, there are other ways in which this method can be used for ligand discovery. In particular, this method can be used to define a relatively narrow region of chemical space, and a combinatorial library can be designed to search that limited space particularly thoroughly. Given the finite synthetic capacity of even the most ambitious combinatorial chemistry schemes, this mechanism can channel synthetic diversity into higher probability directions.

Again, there are several alternative implementations for this computational method. One implementation begins with detailed grids of point charges fit to the optimal multipoles and segregates the grid into regions of space corresponding to pockets appropriate for receiving one or more functional groups. The shape and point charges are then used to assign the general size and character, e.g., positively charged, negatively charged, highly polar, moderately polar, weakly polar, or hydrophobic. These property definitions may be used to bias combinatorial synthesis towards generating appropriate ligands.

Having described the computational aspects of the present computer-implemented process some biological model systems will now be described.

Biological Model Systems

EXAMPLE 1

Class II Major Histocompatibility Complex (MHC) Proteins

Introduction

The major histocompatibility complex proteins (MHC) are cell-surface antigen presenting structures whose role is to display a sample of proteolized intracellular peptide to T cells. Recognition of a peptide as "foreign" by a T cell induces an immune response. This response includes killing the antigen-presenting cell (class MHC, usually) or secreting lymphokines that control attack by various elements of the immune system, including B cell activation (class II MHC, usually). Because each individual has a limited number of histocompatibility proteins and a virtually unlimited number of peptides to present, each MHC molecule is capable of presenting a wide variety of peptides. Structural studies have revealed separate mechanisms used by class I and class II MHC molecules for achieving high affinity yet fairly low specificity (L. J. Stem and D. C. Wiley, Structure 2:245 (1994)).

The structure of the HLA-DR1 class II MHC protein complexed with a peptide from influenza virus has been solved (L. J. Stern, et al., Nature (London) 368:215 (1994)). The original structure of HLA-DR1 in complex with influenza hemagglutinin residues 306–318 (PKYVKQNTLKLAT) elucidated a number of important features of binding and recognition that have been confirmed in other class II MHC complexes. The protein is comprised of an eight-stranded beta-sheet with two immunoglobulin-like domains on the cell-surface side and a pair of alpha-helices on the extracellular side. The peptide-binding site is

a cleft between the two helices and supported by the beta-sheet. The peptide binds in an extended but highly twisted conformation, similar to the type II polyproline helical conformation; the N- and C-termini extend outside of the site. Most of the hydrogen bonds between peptide and protein (12 of 15) are to peptide backbone groups, which helps to explain how the protein recognizes many different peptides. The observed peptide conformation forces each peptide side chain into one of three directions: 5 of the side chains (Y308, Q311, T313, L314, and L316) are directed into pockets in the surface of the MHC molecule and are essentially buried, 6 of the side chains (K307, V309, K310, N312, K315, and T318) are directed out away from the binding site and toward the T cell receptor, and the remaining 2 side chains (P306 and A317) are directed across the site. Thus, residues making extensive contact with the MHC are, for the most part, distinct from those poised to interact with approaching T cell receptors. Of the 5 pockets, the deepest accommodates Tyr308, though binding studies shown that tyrosine, phenylalanine, or tryptophan are all allowed. Different class II MHC alleles incorporate substitutions at the 5 pockets that receive the 5 buried side chains. It is thought that alterations in these interactions are responsible for allotypic differences in peptide specificity. Because individuals differ in their allotypic complement of MHC molecules, individuals differ in the profile of their immune response.

The relative affinity of peptides for binding to individual class II MHC molecules is thought to be responsible for relative peptide antigenicity. Phage display selection and amplification studies have defined the frequency with which each amino acid is found at individual positions in high-affinity peptides (J. Hammer, et al., J. Exp. Med. 176:1007 (1992) and J. Hammer, et al., PNAS U.S.A. 91: 4456 (1994)). The strongest anchor position was determined to be a large aromatic at position 1, which was found as Tyr (48%), Phe (25%), or Trp (13%) predominantly. Position 4 was a long hydrophobe, found as Met (50%) and Leu (28%); position 6 was a small residue, found as Ala (32%) and Gly (22%); and position 9 was generally found as Leu (45%).(J. Hammer, et al., J. Exp. Med. 176:1007 (1992)). Also, there were very few negatively charged side chains recovered at any position. This data provides a useful semi-quantitative set of relative affinities that are useful for validating the computational methodology of the present process.

Testing and Validation

The class II MHC HLA-DR1 system is used to test the computational methodology disclosed herein to analyze the peptide-binding site, and to design enhanced-binding molecules. Testing and validation consists of a number of tasks, initially using the crystal structure with bound viral peptide (L. J. Stem, et al., Nature (London) 368:215 (1994)). These tests are designed to confirm that the methods are capable of (i) recognizing that the observed bound peptide is a good binder, (ii) recognizing that known deleterious peptide mutations are unfavorable, (iii) recognizing that known enhanced-binding peptide mutations are favorable, (iv) reproducing the known pattern of binding hydrophobic, polar, and charged residues in individual surface pockets, and (v) regenerating the known peptide backbone conformation and contacts.

Analysis of Bound-Peptide Complex

The binding of viral peptide to HLA-DR1 is analyzed using the methods disclosed herein. Briefly, the strategies disclosed herein are used to regenerate the bound peptide's charge distribution. A variable point charge is placed on each atom of the peptide and the charge values are computed that

optimize the free energy of binding. By comparing these point charges to the actual point charges, the reduction in affinity of the peptide compared to the calculated optimum can be computed. Discrepancies from point charges between the calculation and the actual point charges of the viral peptide suggest that the possibility exists to design an enhanced-binding ligand. In this manner, this set of tests is used to confirm the asserted utility of the claimed methods with respect to designing an enhanced-binding ligand based upon the structure of a known ligand and its binding partner.

Enhanced- and Reduced-Binding Mutations

Tests of relative affinity are performed initially using isosteric or near-isosteric replacements. From the data of Hammer et al. using phage display studies, Tyr is preferred over Phe at position 1, and Met or Leu is preferred over Gin at position 4, where the underlined residue corresponds that in the bound peptide structure (L. J. Stem, et al., Nature (London) 368:215 (1994); J. Hammer, et al., J. Exp. Med. 176:1007 (1992); and J. Hammer, et al., PNAS U.S.A. 91:4456 (1994)). The methods disclosed herein are used to compute the change in affinity due to these mutations.

The novel strategy disclosed herein for ligand design is to start with a given conformation of receptor (or other target molecule, such as an antibody or an enzyme) and find the properties of the ligand that will optimally complement that conformation. The tests performed described herein assay whether the methods can detect differences in affinity due to differences in the ligand charge distribution, an essential prerequisite for defining the optimal ligand charge distribution. When the point-charge magnitudes are optimized as described in the previous paragraph, it is expected that the polarity assigned to the Tyr1 hydroxyl remains, that of Val2 increase, and that of Gln4 and Thr6 decrease, reflecting the electrostatic tendencies of Hammer et al. (J. Hammer, et al., J. Exp. Med. 176:1007 (1992); and J. Hammer, et al., PNAS U.S.A. 91:4456 (1994)).

Pattern of Polar and Non-Polar Side Chains

The methods of the present computer-implemented process are used to probe the peptide binding site without reference to known positions of peptide atoms. This probing is done in two modes. In one mode, each of the five major binding pockets is probed through individual optimization of the charge distribution in that site only; in the second, the five major binding pockets are probed together, with the charge distribution for the entire site optimized in one computation. A comparison of the results indicates the extent to which the sites are coupled; experimental work suggests that the coupling should be minimal (J. Hammer, et al., J. Exp. Med. 180:2353 (1994)). A complementary shaped region is constructed through sphere packing and the multipolar charge distribution that optimizes binding to the site is computed. Both through direct examination of the multipoles and by constructing a gridded point-charge distribution complementary to the site, each site is categorized as to how well it accepts hydrophobic, polar, positive, and negative groups. Examination may reveal mixed character, where a site is largely hydrophobic but accepting of some localized polar groups (presumably the Tyr1 site is of this type). Comparison with the known site characteristics is done to evaluate the results. A discrepancy may result if the peptide desolvation penalty used in the computation (that which would result from a rigid peptide in the bound conformation) were substantially different from that experienced by actual ligands in phage-display studies. However, we do not anticipate this discrepancy to be of concern because the desolvation penalty is dominated by polar and charged groups, which should be exposed to solvent in the

unbound state and which should be in the observed extended and twisted conformation.

Backbone Trace and Contacts

Because the backbone trace is thought to be invariant for essentially all peptides that bind, one expects the site to strongly dictate backbone contacts. Accordingly, the methodology of the present computer-implemented process is used to regenerate the position of the backbone observed crystallographically to further validate the methods disclosed herein. Using the above-described methods, the distributed multipole description of the optimal ligand in the region of peptide backbone binding is identified, converted to a gridded point-charge field, and the peptide amide groups (N-methyl acetamide) are fitted into the charge field as a least-squares fit while also not allowing steric overlap with the walls of the site.

Design for Enhanced Binding

In general terms, the strategy for designing enhanced binding ligands is used to locate opportunities where known ligands do not take full advantage of the site. To this end, both individual chemical groups that pay more in desolvation energy than they recover in favorable interactions and also sites where current liganding groups fall short of computed optima are identified. The computations carried out above (Testing and Validation) are re-analyzed in search of such opportunities.

Analysis of Bound-Peptide Complex

The complete electrostatic dissection described above is used to detect functional groups (side chains or backbone dipolar groups on both the peptide and the binding site) whose total electrostatic contribution to binding is unfavorable (that is, whose mutation to a hydrophobic group is computed to lead to tighter binding). This electrostatic dissection suggests targeting regions of the peptide (even if they are backbone) for modification to hydrophobic groups to produce a more stable complex. Using this strategy we were able to identify three stabilizing mutations in a variant of the Arc repressor (Z. S. Hendsch, et al., Biochemistry 35:7621 (1996)). An MHC protein group that contributes unfavorably to binding can be ameliorated by modifying the peptide to make improved interactions with it. These opportunities can be confirmed by a number of parallel studies, including the computation in which the point charges of the viral peptide atoms are re-optimized (see above). The same locations for reduction and increase in the polarity of the ligand should be found. Such parallel confirmation is used to provide further evidence that a proposed site can be modified to enhance binding.

Pattern of Polar and Non-Polar Side Chains

It is anticipated that the individual pocket optimizations as well as that of the whole site can be used as the source of suggested detailed changes for the purpose of identifying ligands with enhanced binding to its target molecule. In choosing locations for such optimization, regions where the largest free energy gains can be recovered as measured by discrepancy between actual and optimized charge distribution and corresponding binding energy are initially selected. Three such regions include: the position 1 binding pocket, the peptide-backbone binding area (see below), and a pocket occupied by a solvent cluster in the viral-peptide study (L. J. Stem, et al., Nature (London) 368:215 (1994)). Position 1 accommodates a Tyr in the viral-peptide complex (L. J. Stem, et al., Nature (London) 368:215 (1994)) but is frequently found as Phe or Trp as well (J. Hammer, et al., J. Exp. Med. 176:1007 (1992); and J. Hammer, et al., PNAS U.S.A. 91:4456 (1994). In a recently determined crystal structure of HLA-DR1 with a different peptide bound, Trp

occupies the site. The orientation of the Trp side chain is roughly 90° rotated with respect to the Tyr, yet the surrounding protein pocket is essentially unchanged. Placing a large hydrophobic side chain in this pocket appears virtually a requirement for binding J. Hammer, et al., PNAS U.S.A. 91:4456 (1994)). The optimized charge distribution generated for groups binding to this pocket can be used as a guide to a combinatorial synthetic scheme to synthesize enhanced binding ligands.

Backbone Trace and Contacts

The present computer-implemented process can be used to design ligands having non-peptide backbones for improved binding. By comparing the optimized charge distributions for the backbone-binding region to the peptide charge distribution, improved backbone chemistries can be rationally designed. For example, the method can be used to identify ligands that have the equivalent of an alpha-carbon (or at least a beta-carbon) so permit attachment of the presented side chains onto the T-cell side of any new platform is designed.

EXAMPLE 2

HIV Protease

Introduction

The protease from HIV is required for proper assembly of virus. Inactivation of the protease by mutation leads to the production of non-infectious particles. Design of HIV-protease inhibitors has been a major effort of a number of pharmaceutical companies for the past decade or more. This research was aided by the facility with which high-resolution X-ray crystallographic data could be obtained after proper conditions were worked out for expression, purification, and crystallization. In the Protein Data Bank there are currently over 45 structures of HIV-1 protease either alone or in complex with ligand. These structures provide a rich data set for examining modes of interaction of different ligands with a common protein. A number of very promising inhibitors have already been developed, some are in clinical trials, and a few have been approved by the FDA. Nevertheless, "escape" mutants of the protease have been isolated for a number of these inhibitors.

The protease structure reveals an essentially symmetric homodimer of a 99-residue polypeptide chain. The active site is formed at the two-fold axis, is enclosed by a pair of symmetry-related loops that appear highly flexible in the unbound state but close over the active site upon ligand binding, and adjoins a cleft that can bind substrates up to seven residues long. The active site contains the triad Asp25, Thr26, and Gly27 from each subunit, with the pair of Asp25 carboxylate groups in close proximity and nearly coplanar. The apparent exact two-fold symmetry of the enzyme in the absence of ligand is disrupted somewhat by binding (asymmetric) peptide ligands. One particularly interesting issue in design studies has been whether asymmetric ligands (such as those modeled on peptide substrates) or symmetric ligands (which have the opportunity to bind with the ligand two-fold axis coincident with the enzyme two-fold) are tighter binding. A surprising result has been that certain symmetric ligands are found to bind asymmetrically in the active site. The computational methods disclosed herein can be used to investigate the energetic contributions to this difference.

Substrate specificity studies have been used to determine binding preferences for peptides. These have revealed affinity for Gln or Glu at the P2' position and a largely hydrophobic side chain (Phe, Leu, Met, Asn, or Tyr) at P1. Less pronounced preferences include Glu at P3 and a hydrophobe

at P2 (A. Wlodawer and J. W. Erickson, Annu. Rev. Biochem. 62:543 (1993)).

The application of the methods disclosed herein to the analysis of binding modes of molecules whose involvement is critical to HIV infection can be used to facilitate the design of tight-binding ligand molecules for use as diagnostic and therapeutic agents. In general, the methods of the present computer-implemented process are primarily continuum electrostatics and secondarily free energy simulations. The process provides a novel method for finding the electrostatic complement of a target molecule. The preliminary results demonstrate that the computational modeling used herein for molecular and energetic dissection for a continuum analysis yield conclusions that are consistent with those found in by using a more detailed (and time-consuming) free energy simulation for a pair of studies on protein-DNA recognition by 434 repressor (see, example 3).

Testing

Initial testing of the fundamentals of the method are carried out in studies of the class II MHC molecule, and next carried out using the HIV-1 protease. Accordingly, the above-described methods are used to design enhanced-binding ligands that bind to the HIV-1 protease. One difficulty encountered with many ligand design protocols is the need to predict the conformation of bound complexes. The present computer-implemented process circumvents this problem by choosing a conformation of the protein and solving for a set of molecular descriptors for an optimally complementary ligand. The process also provides tools to examine a subset of the available structures of HIV-1 protease, both alone and in complex with various ligands.

Loop Conformation

Two symmetry related loops are in an open conformation in the unbound form of the enzyme and close down against the active site in the bound form. One set of inhibitors that is well characterized and is attractive due to their relative rigidity is the cyclic urea compounds being developed by DuPont Merck Pharmaceuticals (P. Y. S. Lam, et al., Science 263:380 (1994) and C. N. Hodge, et al., Chem. & Biol. 3:301 (1996)). Members of this family of compounds can be used to analyze the bound-state structure. For example, the complex with XK 263 (a symmetric cyclic urea with two naphthyl, two phenyl, and two hydroxyl substituents) is in the Protein Data Bank, and the complexes with DMP 323 and DMP 450 are shown in FIG. 4 ("Inhibitors of HIV-1 Protease"). The effect of this conformational change is examined on the computed properties of the optimal ligand using the above-described methods. The effect of receptor conformational change on complementary ligand properties is examined by characterizing the optima by the shapes and relative polarities of the moieties occupying individual subsite pockets at the active site. It is anticipated that the differences in computations will be rather small, since the substrate must initiate binding with the loops in the open conformation and complete binding when the loops are closed. Substrates either represent some compromise between being complementary to the open and closed form, or there isn't substantial difference between the two.

Protonation State

One important and currently unresolved question central to the design of protease inhibitors is the protonation state of the catalytic aspartyl residues (Asp 25 and 25'). It is anticipated that the protonation state of this pair of side chains will substantially change the properties of the computed electrostatic complement. It is potentially worthwhile for a ligand to incur greater desolvation penalty to interact with a charged, rather than an uncharged, aspartic acid. It is anticipated that a comparison of the computed optimal ligand electrostatic properties to actual bound ligands will permit the assignment of protonation states to some of these complexes. The case of the cyclic ureas from the DuPont Merck group are useful in this study because NMR evidence is consistent with the aspartyl groups each being protonated (D. A.

Torchia, et al., J. Am. Chem. Soc. 116:1149 (1989)). By comparing the optimal complements from computations using a doubly-, singly-, and unprotonated catalytic pair, the affect of the availability of chemical freedom in an active site on its ligand binding properties is determined. Such studies also permit the identification of a preferred titration state that is more susceptible to ligand binding than others.

Symmetric and Asymmetric Binding

A number of symmetric inhibitors have been designed based on the principle that they would be more complementary to the symmetric enzyme (M. Miller, et al., Science 246:1149 (1989)). Although some of these have been observed to binding symmetrically in crystallographic studies (XK 263 (P. Y. S. Lam, et al., Science 263:380 (1994), DMP 450 (C. N. Hodge, et al., Chem. & Biol. 3:301 (1996)), and A-76928 (M. V. Hosur, et al., J. Am. Chem. Soc. 116:847 (1994)), others bind asymmetrically (A-76889 (M. V. Hosur, et al., J. Am. Chem. Soc. 116:847 (1994)). There could be two reasons for asymmetric binding of a symmetric inhibitor. Either the site can deform so that it is truly complementary to an asymmetric ligand, or the site can remain essentially symmetric but the ligand preferentially makes asymmetric interactions. These cases can be distinguished more precisely by examining the computed electrostatic complement for sites harboring symmetrically and asymmetrically bound ligands for two-fold symmetry. If the complement remains symmetric for asymmetrically bound ligands, improvements to the ligand can be defined using the above-described methods. For example, enhanced-binding ligands can be designed by studying the four compounds in FIG. 4 which represent different choices (both symmetric and asymmetric) for using hydroxyl groups to compensate the buried catalytic aspartic acid side chains.

Design

The approaches to design of protease inhibitors are similar to those described above in reference to the design of MHC ligands. A few design points unique to HIV protease are described herein.

Each of the above-described studies answers specific questions about how conformational and titration changes to active sites affect the properties of the computed complementary ligand. Each study also can be analyzed for opportunities to modify existing ligands (to obtain "enhanced-binding ligands") or to design entirely new ligands with enhanced affinity. Alcohols and diols are prevalent in a number of HIV-1 protease inhibitors; more effective moieties to satisfy the electrostatic properties of the aspartyl groups can be identified to design enhanced-binding ligands.

One particularly important problem with all drugs targeted to HIV is the eventual evolution of "escape" mutants.

The invention is useful for developing the minimal charge configuration required to complement the active site residues. It is believed that such a core molecule is useful because its limited size should reduce the number of contacts potentially disruptable by escape mutants. In addition, since contacts would all be at the catalytic site, disrupting mutants could be inactive.

Design Studies on Other HIV Targets

Design studies against other HIV targets are performed using the above-described methods. Other HIV targets include the RNA complexes of TAR and RRE and the HIV envelope proteins.

EXAMPLE 3

The 434 Repressor DNA-binding Domain

Introduction

We have analyzed the high-resolution X-ray crystal structure of the 434 repressor DNA-binding domain, R1-69, bound to the O_R1 operator using continuum electrostatic calculations. The principal results are outlined below. The interaction was dissected into contributions from each protein backbone carbonyl, $C_\alpha NH$, and side chain and each nucleic-acid ribose, base, and phosphate. For each group a desolvation contribution to binding was calculated as well as contributions from new interactions made across the interface (termed "intermolecular") and from changes in the screening within the protein or DNA (termed "intramolecular" interactions).

Currently only the rigid binding of pre-conformed protein dimer to pre-conforned DNA has been studied. These methods can be extended to address conformational flexibility as described above. The overall electrostatic contribution to binding is unfavorable (45.3 kcal/mol). This is due to a large desolvation penalty (132.9 kcal/mol) that is only partially offset by favorable intermolecular terms (−96.4 kcal/mol). The sum of intramolecular terms is small and unfavorable (8.8 kcal/mol). Four salt bridges formed in the complex (two symmetry-related pairs) stabilize complex formation by an average of −1.7 kcal/mol each. This is due largely to the fact that these groups incur a smaller desolvation penalty than do protein side chains in folding from the unfolded state. In this regard, binding appears to be somewhat different from folding, but our further results show that the distinction is somewhat more complex.

The largest contributors to the desolvation penalty come from the charged groups in the system—protein side chains (63.5 kcal/mol) and DNA backbone groups (50.6 kcal/mol). Protein backbone groups (6.9 kcal/mol) and DNA bases (11.9 kcal/mol) incur much smaller costs. The desolvation penalty is substantial for many groups that become buried at the protein-DNA interface. Interestingly, some side chains that lie nearby but not at the interface also lose significant solvation on binding.

The strong, favorable intermolecular interactions formed in the complex are made almost entirely with DNA backbone groups. Surprisingly, equal amounts come from interactions with protein side chains (−42.2 kcal/mol), which include a large number of charged groups, as with the protein backbone (−41.8 kcal/mol), which is only polar except for the charged termini. The analysis points to strong interactions made between the N-termini of alpha-helices

and individual phosphate groups as well as the strong interaction between the backbone groups in a protein loop with a set of phosphates. Additionally, some interactions between charged side chains and the DNA backbone are rather distant, but are directed through the low-dielectric protein, where electrostatic interactions might be expected to be longer range due to less screening by solvent.

Intermolecular interactions with the bases are small: 2.2 kcal/mol (unfavorable) with protein backbone and −14.6 kcal/mol (favorable) with protein side chains, although the base-side chain interactions are generally thought to confer substantial specificity to protein-DNA complexes. Interestingly, interactions close enough to make a hydrogen bond account for roughly half the favorable intermolecular interactions; an equal contribution comes from interactions too distant to be hydrogen bonding. In particular, many of these more distant interactions are to the "non-contacted" bases in the central region of the operator; Arg43 in the left and right half-sites contribute −3.9 kcal/mol.

Overall the intramolecular interactions contribute only 8.8 kcal/mol to the electrostatic docking energy, but the sources of this effect are quite interesting. These interactions exist in the identical geometry in the bound and free states since they are within the protein or DNA. Their magnitude changes on docking, however, because the removal of high-dielectric solvent in the complex reduces the screening of interactions. Repulsions within the DNA backbone (due largely to phosphate-phosphate interactions) increase in magnitude by 19.2 kcal/mol on binding protein because the reduced solvent screening in the bound state leads to a lower effective dielectric. This is partially offset by a favorable contribution between protein side chains of −11.9 kcal/mol, which is due largely to attractive salt bridges within the protein whose strength "increases" due to reduced screening in the complex.

When all of the contributions (desolvation, intermolecular, and intramolecular) are tabulated for each group, most groups individually pay more in desolvation energy than they recover in other interactions. This is particularly true for the phosphate groups and all but one base, as well as for the side chains at the binding interface. Groups that do recover more than they pay in desolvation energy tend to be largely buried in the undocked state.

In summary, this work demonstrates the detailed insights that result from an energetic dissection of a binding event. These techniques are useful for exploring ligand binding to HIV targets, and permit the rational design of enhanced-binding ligands.

Free Energy Analvsis of the Effect of a Point Mutation: Simulation of a Base-Pair Change in a 434 Repressor-DNA Complex

To address specific issues of recognition and to validate the results of our continuum electrostatic investigation, we carried out a free energy simulation study with explicit solvent. The bound-state starting structure was the high-resolution complex of R1-69 bound to O_R2 (L. J. W. Shimon and S. C. Harrison, J. Mol. Biol. 232:826–838 (1993)). The mutation was TA→GC at position 7L. Multiple unbound conformations of the DNA were generated from a 300-ps molecular dynamics trajectory. Five frames of the trajectory were chosen and used as starting structures for the unbound-state free energy calculation. Although most of the interac-

tions between 434 repressor and DNA are in the major groove, this operator mutation occurs near the pseudo-dyad axis where repressor faces the minor-groove side of DNA. A total of ten simulations were carried out in the unbound state and six in the bound state, which led to good statistics. The results are outlined briefly here (E. J. Simon, "A Molecular Dynamics Study of a Mutation in a Bacteriophage 434 Operator/Repressor Complex", PhD thesis, Harvard University (1996)).

Results:

The overall stability change is +1.4±0.7 kcal/mol, which disfavors binding to the mutant operator. This is in good agreement with experimental values of 0.8–1.2 kcal/mol (G. B. Koudelka, et al., Nature (London) 326:886 (1987)). An analysis of the source of this overall stability change (B. Tidor, "Molecular Modeling of Contributions to Free Energy Changes: Applications to Proteins. PhD thesis, Harvard University (1990); B. Tidor and M. Karplus, Biochemistry 30:3217 (1991); and B. Tidor Proteins: Struct., Funct., Genet. 19:310 (1994)) was carried out and shows a strong repulsion between the side chain of Arg43L and the N2 amino group of the mutant guanine. This is a remarkably interesting interaction because it suggests that this arginine acts as a negative determinant of specificity by "interfering" with a guanine at this position. The array of hydrogen-bond acceptors in the minor groove in this AT-rich region of the mutation site and the flanking phosphate groups polarize the surrounding solvent water to interact favorably with this negative potential. The introduction of the Gua N2 donor to the minor groove effectively repels this polarized solvent. The repulsion is stronger in the unbound than the bound state because solvent is displaced from this region of the minor groove on protein binding.

Comparison of these free energy simulation results with the continuum electrostatic study shows essentially the same dissection for the interactions of Arg 43, including the solvent polarization effect. This comparison demonstrates the accuracy of continuum methodology relative to explicit simulations. The present computer-implemented process is based around the continuum approach, which is more economical and can be used to analyze an entire binding site at once, rather than one group at a time. Free energy simulations are used primarily to examine points of disagreement between continuum theory and experiment.

APPENDIX

FIG. 5. illustrates problem geometries. FIG. 5*a*. shows the binding reaction between a receptor (R) and spherical ligand (L) that dock rigidly to form a spherical bound-state complex. Receptor, ligand, and complex are all low-dielectric media ($\in_1$) that are surrounded by high-dielectric solvent ($\in_2$). FIG. 5*b*. shows that the boundary-value problem solved here involves a charge distribution in a spherical region of radius R with dielectric constant $\in_1$ surrounded by solvent with dielectric constant $\in_2$. The origin of coordinates is the center of the larger spherical region, but the charge distribution is expanded in multipoles about a point a distance d along the z-axis. The geometric requirement is that the ligand sphere not extend beyond the receptor sphere, $R \geqq d+a$, although the case of equality is illustrated in the figure.

The electrostatic free energy of binding is the difference between the electrostatic free energy in the bound and the unbound state, $\Delta G_{binding}=G^{bound}-G^{unbound}$ (see FIG. 5*a*). Because the dielectric model includes responses that affect the entropy as well as the en-thalpy, the electrostatic energy is considered to be a free energy. The free energy of each state is expressed as a sum of coulombic and reaction-field (hydration) terms involving the ligand (L), the receptor (R), and their interaction (L–R):

$$G^{state}=G_{coul,L}{}^{state}+G_{coul,R}{}^{state}+G_{coul,L-R}{}^{state}+G_{hyd,L}{}^{state}+G_{hyd,R}{}^{state}+G_{hyd,L-R}{}^{state}. \quad (1)$$

This results in the following expression for the binding free energy,

$$\Delta G_{binding}=\Delta G_{coul,L-R}+\Delta G_{hyd,L-R}+\Delta G_{hyd,L}+\Delta G_{hyd,R} \quad (2)$$

where the fact that the geometry of point charges in the receptor and ligand remain fixed is used in the model to cancel the coulombic self contribution of ligand and receptor and where the two L–R terms are due only to the bound state because the ligand and receptor are assumed not to interact in the unbound state. (Note, however, that the charge distribution for the receptor need not be the same in the bound and unbound states. If they are different, this adds a constant to $\Delta G_{binding}$ that can be dropped in defining ΔG_{var} in Eq. (3)). Thus, Eq. (2) describes the electrostatic binding free energy as a sum of desolvation contributions of the ligand and the receptor (which are unfavorable) and solvent-screened electrostatic interaction in the bound state (which is usually favorable). Since the goal is to vary the ligand charge distribution to optimize the electrostatic binding free energy and the last term simply adds a constant, a relevant variational binding energy is defined,

$$\Delta G_{var}=\Delta G_{int,L-R}+\Delta G_{hyd,L} \quad (3)$$

in which the first two terms on the right hand side (RHS) of Eq. (2) have been combined into a screened interaction term and the constant term has been dropped. Note that

$$\Delta G_{int,L-R}=\sum_{j\in R} q_j V_L^{bound}(\vec{r_j})=\sum_{j\in R} q_j[V_{coul,L}^{bound}(\vec{r_j})+V_{hyd,L}^{bound}(\vec{r_j})] \quad (4)$$

and

$$\Delta G_{hyd,L}=\frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{bound}(\vec{r_i})-\frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{unbound}(\vec{r_i}) \quad (5)$$

where $V_L{}^{state}$ is the total electrostatic potential in the indicated state due to the ligand charge distribution only and $V_{term,L}{}^{state}$ is the coulombic or reaction-field (hydration) term, as indicated. The summations are over atomic point charges in the ligand ($i\in L$) or receptor ($j\in R$). The factor of ½ in Eq. (5) is due to the fact that the ligand charge distribution interacts with the self-induced reaction field.

$V_{coul,L}{}^{bound}$, $V_{hyd,L}{}^{bound}$, and $V_{hyd,L}{}^{unbound}$, the three electrostatic potentials in Eqs. (4) and (5), are expressed in terms of the given geometry and charge distribution by solving the boundary-value problem shown in FIG. 5*b*. A charge distribution (corresponding to the ligand) is embedded in a sphere of radius R. The center of the sphere is taken as the origin of coordinates (unprimed) but the charge distribution in multipoles is expanded about a second origin (primed) translated a distance d along the z-axis, so that

$$\vec{r}(r,\theta,\phi)=\vec{d}(d,\theta_d=0,\phi_d=0)+\vec{r}'(r',\theta',\phi'). \quad (6)$$

The potential everywhere satisfies the Poisson equation. Inside the sphere, it may be written as,

$$V_{in}(\vec{r}) = \sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r_i}|} + \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{l,m} r^l Y_{l,m}(\theta, \phi) \quad (7)$$

where the first term on the RHS is the coulombic and the second is the reaction-field (hydration) potential, and the summation over i corresponds to the ligand point charges. Outside the sphere, the coulombic and reaction-field potential can be combined and written as,

$$V_{out}(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{B_{l,m}}{r^{l+1}} Y_{l,m}(\theta, \phi) \quad (8)$$

where $A_{l,m}$ and $B_{l,m}$ are to be determined by the proper boundary conditions and $Y_{l,m}(\theta, \phi)$ are the spherical harmonics. The coulombic term in Eq. (7) is expanded in spherical harmonics and multipoles of the charge distribution about the center of the sphere. Here the origin of the multipole expansion is shifted to $\vec{d}$,

$$\sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r_i}|} = \sum_i \frac{q_i}{\epsilon_1 |(\vec{r} - \vec{d}) - \vec{r_i}|} = \sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r_i}|} \quad (9)$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} Q'^*_{l,m} \frac{Y_{l,m}(\theta', \phi')}{\epsilon_1 r'^{l+1}} \quad (10)$$

where $Q'_{l,m}$ is a spherical multipole expanded about the primed origin, $\vec{d}$,

$$Q'^*_{l,m} = \sum_i q_i r_i'^l Y^*_{l,m}(\theta_i', \phi_i'). \quad (11)$$

The definition of the $Y_{l,m}(\theta, \phi)$ used by Jackson[1] is adopted. The expression in Eq. (10) is valid for $r'>r'_i$ (i.e., outside the ligand or, more precisely, outside the sphere whose center is at $\vec{d}$ and whose radius is the longest distance between $\vec{d}$ and a point charge).

To substitute into Eq. (7) and combine terms involving spherical harmonics, first $Y_{l,m}(\theta', \phi')/r'^{l+1}$ of Eq. (10) is expanded in terms of $Y_{l,m}(\theta, \phi)/r^{l+1}$. This is done using the results of Greengard,[2] which state that for r>d,

$$\frac{Y_{l,m}(\theta', \phi')}{r'^{l+1}} = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} K_{l',m',l,m} \left[\frac{4\pi(2l+1)}{(2l'+1)(2l'+2l+1)}\right]^{\frac{1}{2}} d^{l'} Y^*_{l',m'}(\theta_d, \phi_d) \frac{Y_{l'+l,m'+m}(\theta, \phi)}{r^{l'+l+1}} \quad (12)$$

where

$$K_{l',m',l,m} = \left[\frac{(l'+l+m'+m)!(l'+l-m'-m)!}{(l'+m')!(l'-m')!(l+m)!(l-m)!}\right]^{\frac{1}{2}}. \quad (13)$$

Since a geometry with θ_d=0 (FIG. 1*b*) has been used, only m'=0 terms in Eq. (12) are non-vanishing, in which case Eq. (10) becomes,

$$\sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r_i}|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{\epsilon_1} \left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} Q'^*_{l,m} \sum_{l'=0}^{\infty} K_{l',0,l,m} d^{l'} \left(\frac{4\pi}{2l'+2l+1}\right)^{\frac{1}{2}} \frac{Y_{l'+l,m}(\theta, \phi)}{r^{l'+l+1}} \quad (14)$$

in which the multipole distribution is taken about the point $\vec{d}$, but the potential is expressed as a summation of spherical harmonics about the large-sphere center. The above equation can also be written as,

$$\sum_i \frac{q_i}{\epsilon_1 |\vec{r} - \vec{r_i}|} = \quad (15)$$

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{\epsilon_1} \left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} \frac{Y_{l,m}(\theta, \phi)}{r^{l+1}} \sum_{l'=|m|}^{l} K_{l-l',0,l',m} d^{l-l'} \left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} Q'^*_{l',m}$$

where terms with the same $Y_{l,m}(\theta, \phi)$ are grouped together, as opposed to Eq. (14), where terms with the same $Q'^*_{l,m}$ are grouped.

Upon substituting Eq. (15) into Eq. (7) and matching boundary conditions at r=R,

$$V_{in}|_{r=R} = V_{out}|_{r=R} \quad (16)$$

$$\epsilon_1 \frac{\partial V_{in}}{\partial r}\bigg|_{r=R} = \epsilon_2 \frac{\partial V_{out}}{\partial r}\bigg|_{r=R} \quad (17)$$

the hydration (reaction-field) potential inside the sphere is,

$$V_{hyd}(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{l,m} r^l Y_{l,m}(\theta, \phi) \quad (18)$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} r^l Y_{l,m}(\theta, \phi) \left(\frac{C_l}{R^{2l+1}}\right)$$

$$\sum_{l'=|m|}^{l} K_{l-l',0,l',m} d^{l-l'} \left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} Q'^*_{l',m} \quad (19)$$

where

$$C_l = \frac{(\epsilon_1 - \epsilon_2)}{\epsilon_1 [\epsilon_2 + l\epsilon_1/(l+1)]}. \quad (20)$$

The various V's can he rewritten, with their dependence on the $Q'^*_{l,m}$ made explicit. $V'_{coul,L}{}^{bound}$ is given by Eq. (10), $V_{hyd,L}{}^{bound}$ is given by Eq. (19) but rewritten so that the terms with the same $Q'^*_{l,m}$ are collected, and $V_{hyd,L}{}^{unbound}$ is given by Eq. (19) with R=a and d=0.

$$V'^{bound}_{coul,L}(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} Q'^*_{l,m} \frac{Y_{l,m}(\theta', \phi')}{\epsilon_1 r'^{l+1}} \quad (21)$$

-continued

$$V_{hyd,L}^{bound}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} \tag{22}$$

$$\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} Q_{l,m}^{l'*} K_{l'-l,0,l,m} d^{l'-l}\left(\frac{C_{l'}}{R^{2l'+1}}\right) r^{l'} Y_{l',m}(\theta,\phi)$$

$$V_{hyd,L}^{unbound}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\left(\frac{C_l}{\alpha^{2l+1}}\right) Q_{l,m}^{l*} r^l Y_{l,m}(\theta,\phi) \tag{23}$$

Substituting into Eq. (4), the dependencc of $\Delta G_{int,L-R}$ on the $Q'^*_{l,m}$ is made explicit,

$$\Delta G_{int,L-R} = \sum_{j\in R} q_j\left[V_{coul,L}^{ibound}(\vec{r_j}) + V_{hyd,L}^{bound}(\vec{r_j})\right] \tag{24}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Q_{l,m}^{l*}\sum_{j\in R} q_j\left[\left(\frac{4\pi}{2l+1}\right)\frac{Y_{l,m}(\theta_j',\phi_j')}{\epsilon_1 r_j'^{l+1}} + \right. \tag{25}$$

$$\left.\sum_{l'=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}} K_{l'-l,0,l,m} d^{l'-l}\left(\frac{C_{l'}}{R^{2l'+1}}\right)\right.$$

$$\left. r_j^{l'} Y_{l',m}(\theta_j,\phi_j)\right]$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m} Q_{l,m}^{l*} \tag{26}$$

where in the last line the element $\alpha_{l,m}$ is defined, which is independent of the $Q'^*_{l,m}$, to be the factor multiplying $Q'^*_{l,m}$ in Eq. (25). Each $\alpha_{l,m}$ expresses the contribution of a multipole to $\Delta G_{int,L-R}$ and contains all information concerning the receptor charge distribution required to obtain ΔG_{var}.

For $\Delta G_{hyd,L}$ it is useful to re-express Eq. (5) in terms of the $Q'^*_{l,m}$, the multipoles describing the ligand charge distribution, rather than the individual charges, q_i. $V(\vec{r})$ is expanded around the center of the multipole expansion, $\vec{d}$,

$$\sum_{i\in L} q_i V(\vec{r_i}) = \sum_{i\in L} q_i V\left(\vec{d}+\vec{r_i}\right) \tag{27}$$

$$= \sum_{i\in L} q_i\left[V(\vec{d}) + \vec{r_i}\cdot\vec{\nabla}V(\vec{d}) + \ldots\right]. \tag{28}$$

It has been shown by Rose that in spherical coordinates the expansion becomes,[3]

$$\sum_{i\in L} q_i V\left(\vec{d}+\vec{r_i}\right) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{(2l+1)!!} Q_{l,m}^{l*} y_{l,m}(\vec{\nabla}) V(\vec{d}) \tag{29}$$

where

$$y_{l,m}(\vec{r}) \equiv r^l Y_{l,m}(\theta,\phi) \tag{30}$$

and $y_{l,m}(\vec{\nabla})$ is the operator obtained by replacing $\vec{r}$ with $\vec{\nabla}$. For positive m and when $y_{l,m}(\vec{\nabla})$ operates on a solution of the Laplace equation (i.e., $r^l Y_{l,m}(\theta,\phi)$ or $Y_{l,m}(\theta,\phi)/r^{l+1}$), it has been shown that,[3]

$$y_{l,m}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_1^m \nabla_0^{l-m} \tag{31}$$

for $m \geq 0$.

The double-factorial is defined as

$$(2l+1)!! = (2l+1)\cdot(2l-1)\cdot(2l-3)\ldots 3\cdot 1 \tag{32}$$

$$= \frac{(2l+1)!}{2^l l!} \tag{33}$$

and the spherical partial derivatives are

$$\nabla_1 = -\frac{1}{\sqrt{2}}(\nabla_x + i\nabla_y),\ \nabla_{-1} = \frac{1}{\sqrt{2}}(\nabla_x - i\nabla_y),\ \nabla_0 = \nabla_z. \tag{34}$$

To compute $y_{l,m}(\vec{\nabla})$ for negative m, the fact that $Y_{l,-m}(\theta,\phi) = (-1)^m Y^*_{l,m}(\theta,\phi)$ is used and the definitions of spherical partial derivatives in Eq. (34) to obtain,

$$y_{l,-m}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_{-1}^m \nabla_0^{l-m} \tag{35}$$

for $m \geq 0$.

The hydration energy of the bound ligand is then

$$G_{hyd,L}^{bound} = \frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{bound}\left(\vec{d}+\vec{r_i}\right) = \frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\frac{4\pi}{(2l'+1)!!} Q_{l',m'}^{l'*} y_{l',m'}(\vec{\nabla}) V_{hyd,L}^{bound}(\vec{d}) \tag{36}$$

$$= \frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\frac{4\pi}{(2l'+1)!!} Q_{l',m'}^{l'*} y_{l',m'}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l''=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l''+1}\right)^{\frac{1}{2}} \times \tag{37}$$

$$Q_{l,m}^{l''*} K_{l''-l,0,l,m} d^{l''-l}\left(\frac{C_{l''}}{R^{2l''+1}}\right) y_{l',m'}(\vec{\nabla})\left(r^{l''} Y_{l'',m}(\theta,\phi)\right)\Big|_{\vec{r}=\vec{d}}.$$

To evaluate $y_{l,m}(\vec{\nabla})$ in Eq. (37), Eq. (31) and the gradient formula are used,[4]

$$\vec{\nabla}(\Phi(r)Y_{l,m}(\theta,\phi)) = -\left(\frac{l+1}{2l+1}\right)^{\frac{1}{2}}\left(\frac{d\Phi(r)}{dr} - \frac{l}{r}\Phi(r)\right)T_{l,l+1,m}(\theta,\phi) + \left(\frac{l}{2l+1}\right)^{\frac{1}{2}}\left(\frac{d\Phi(r)}{dr} + \frac{l+1}{r}\Phi(r)\right)T_{l,l-1,m}(\theta,\phi) \quad (38)$$

where

$$T_{l,l',m}(\theta,\phi) \equiv \sum_{m'\in[-1,0,1]} C(l',1,l;m-m',m')Y_{l',m-m'}(\theta,\phi)\hat{\xi}_{m'} \quad (39)$$

the C(l', 1, l; m−m', m') are the vector addition (or Clebsch-Gordon) coefficients frequently encountered in the study of angular momentum shown in Table I,[4] and $\hat{\xi}_m$, are spherical unit vectors,

$$\hat{\xi}_1 = -\frac{1}{\sqrt{2}}(\hat{x}+i\hat{y}),\ \hat{\xi}_{-1} = \frac{1}{\sqrt{2}}(\hat{x}-i\hat{y}),\ \hat{\xi}_0 = \hat{z}. \quad (40)$$

Accordingly,

$$\vec{\nabla} = \hat{x}\nabla_x + \hat{y}\nabla_y + \hat{z}\nabla_z = -\hat{\xi}_1\nabla_{-1} - \hat{\xi}_{-1}\nabla_1 + \hat{\xi}_0\nabla_0. \quad (41)$$

From Eqs. (38) through (41),

$$\nabla_u(r^l Y_{l,m}(\theta,\phi) = (-1)^u[l(2l+1)]^{1/2}C(l-1,1,l;m+u,-u)r^{l-1}Y_{l-1,m+u}(\theta,\phi) \quad (42)$$

Using Table I, Eq. (31), and Eqs. (37) through (42), the following intermediate results are obtained,

$$\nabla_0^{l'-m'}\left(r^{l''}Y_{l'',m}\right) = \left[\frac{(2l''+1)(l''+m)!(l''-m)!}{(2l''-2l'+2m'+1)(l''-m-l'+m')!(l''+m-l'+m')!}\right]^{\frac{1}{2}} r^{l''-l'+m'}Y_{l''-l'+m',m} \quad (43)$$

$$\nabla_1^{m'}\left(r^{l''-l'+m'}Y_{l''-l'+m',m}\right) = (-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''-m-l'+m')!}{2m'(2l''-2l'+1)(l''-m-l'-m')!}\right]^{\frac{1}{2}} r^{l''-l'}Y_{l''-l',m+m'} \quad (44)$$

$$\nabla_{-1}^{m'}\left(r^{l''-l'+m'}Y_{l''-l'+m',m}\right) = (-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''+m-l'+m')!}{2m'(2l''-2l'+1)(l''+m-l'-m')!}\right]^{\frac{1}{2}} r^{l''-l'}Y_{l''-l',m-m'} \quad (45)$$

and the final expression for the hydration energy of the ligand in the bound state,

$$G_{hyd,L}^{bound} = \frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{bound}(\vec{r}_i) = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty} Q_{l,m}^{\prime *}Q'_{l',m}\sum_{l''=\max(l,l')}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}}\frac{C_{l''}}{R^{2l''+1}}\times \quad (46)$$

$$\frac{(l''+m)!(l''-m)!}{(l''-l)!(l''-l')!}\left[\frac{1}{(l+m)!(l-m)!(l'+m)!(l'-m)!}\right]^{\frac{1}{2}} d^{2l''-l-l'}$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\beta_{l,m,l',m'}Q_{l,m}^{\prime *}Q'_{l',m'} \quad (47)$$

where $\beta_{l,m,l',m'}$ is defined by the above two equations; note that $\beta_{l,m,l',m'}$ is zero for m' m.

The hydration energy of the unbound ligand is obtained by setting d=0 and R=α in Eq. (46),

$$G_{hyd,L}^{unbound} = \frac{1}{2}\sum_{i\in L} q_i V_{hyd,L}^{unbound}(\vec{r}_i) = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\left(\frac{C_l}{a^{2l+1}}\right)Q_{l,m}^{\prime *}Q'_{l,m} \quad (48)$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q_{l,m}^{\prime *}Q'_{l,m} \quad (49)$$

where $\gamma_{l,m}$ is defined by Eqs. (48) and (49). Then, $\gamma_{l,m}$ is written as a function of both l and m for notational convenience, although there is no formal dependence on m.

Thus ΔG_{var} has been expressed as a function of the multipoles of the ligand charge distribution, $Q'_{l,m}$ (expanded about the center of the ligand sphere) and the elements $\alpha_{l,m}$, $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ which do not depend on $Q'_{l,m}$. Combining Eqs. (26), (47), and (49) gives

$$\Delta G_{var} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q_{l,m}^{\prime *} + \sum_{l=0}^{\infty}\sum_{m=-1}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\beta_{l,m,l',m'}Q_{l,m}^{\prime *}Q'_{l',m'} - \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q_{l,m}^{\prime *}Q'_{l,m}. \quad (50)$$

Note that only the $\alpha_{l,m}$ depend on the receptor charges, while the $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ depend solely on the geometry of the bound and unbound states. While ΔG_{var}^{opt} is a real quantity, the $\alpha_{l,m}$ and $Q'_{l,m}$ are complex and the products $\alpha_{l,m}Q'^{*}_{l,m}$ and $Q'^{*}_{l,m}Q'_{l',m}$ involve summations over terms of the form $Y^{*}_{l',m'}(\theta',\phi')Y_{l,m}(\theta,\phi))$; note that the $\beta_{l,m,l'm'}$ and $\gamma_{l,m}$ are real. Then ΔG_{var}^{opt} is rewritten in terms of the real and imaginary parts of $\alpha_{l,m}$ and $Q'_{l,m}$,

$$\Delta G_{var} = \sum_{l=0}^{\infty}\left[\alpha_{l,0}Q'_{l,0} + 2\sum_{m=1}^{l}(\mathrm{Re}\alpha_{l,m}\mathrm{Re}Q'_{l,m} + \mathrm{Im}\alpha_{l,m}\mathrm{Im}Q'_{l,m})\right] + \sum_{l=0}^{\infty}\sum_{l'=0}^{\infty}\left[\beta_{l,0,l',0}Q'_{l,0}Q'_{l',0} + 2\sum_{m=1}^{l}\beta_{l,m,l',m}(\mathrm{Re}Q'_{l,m}\mathrm{Re}Q'_{l',m} + \mathrm{Im}Q'_{l,m}\mathrm{Im}Q'_{l',m})\right] - \sum_{l=0}^{\infty}\left[\gamma_{l,0}Q'^{2}_{l,0} + 2\sum_{m=1}^{l}\gamma_{l,m}(\mathrm{Re}Q'^{2}_{l,m} + \mathrm{Im}Q'^{2}_{l,m})\right] \quad (51)$$

(where the summations over m are excluded for l=0) by noting again that $Y_{l,-m}(\theta,\phi) = (-1)^m Y^{*}_{l,m}(\theta,\phi)$ and

$$Y^{*}_{l,m}(\theta',\phi')Y_{l,m}(\theta,\phi) + Y^{*}_{l,-m}(\theta',\phi')Y_{l,-m}(\theta,\phi) = Y^{*}_{l,m}(\theta',\phi')Y_{l,m}(\theta,\phi) + Y_{l,m}(\theta',\phi')Y^{*}_{l,m}(\theta,\phi) \quad (52)$$

$$= 2[\mathrm{Re}Y_{l,m}(\theta',\phi')\cdot\mathrm{Re}Y_{l,m}(\theta,\phi) + \mathrm{Im}Y_{l,m}(\theta',\phi')\cdot\mathrm{Im}Y_{l,m}(\theta,\phi)]. \quad (53)$$

The new variables $ReQ'_{l,m}$ and $ImQ'_{l,m}$ are re-indexed and renamed Q_i as follows,

$$\{Q'_{0,0},Q'_{1,0},ReQ'_{1,1},ImQ'_{1,1},Q'_{2,0},ReQ'_{2,1},ImQ'_{2,1},ReQ'_{2,2},\ldots\}\leftrightarrow\{Q_1,Q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8,\ldots\}. \quad (54)$$

and similar transformations are used to create α_i, β_{ij}, and γ_i. Eq. (51) can then be written as

$$\Delta G_{var} = \sum_{i=1}^{\infty}\alpha_i Q_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\beta_{ij}Q_iQ_j - \sum_{i=1}^{\infty}\gamma_i Q_i^2 \quad (55)$$

$$= \sum_{i=1}^{\infty}\alpha_i Q_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(\beta_{ij}-\delta_{ij}\gamma_i)Q_iQ_j \quad (56)$$

or in matrix notation,

$$\Delta G_{var} = \vec{Q}^T\vec{\vec{B}}\,\vec{Q}+\vec{Q}^T\vec{A} \quad (57)$$

$$=\left(\vec{Q}+\frac{1}{2}\vec{\vec{B}}^{-1}\vec{A}\right)^T\vec{\vec{B}}\left(\vec{Q}+\frac{1}{2}\vec{\vec{B}}^{-1}\vec{A}\right)-\frac{1}{4}\vec{A}^T\vec{\vec{B}}^{-1}\vec{A} \quad (58)$$

where $\vec{Q}$ is the vector formed by the Q_i, $\vec{A}$ is the vector formed by the α_i, $\vec{\vec{B}}$ is the symmetric matrix formed by the $(\beta_{ij}-\delta_{ij}\gamma_i)$, and completion of the square has been used to arrive at Eq. (58). Since $\vec{Q}^T\vec{\vec{B}}\,\vec{Q}$ in Eq. (57) corresponds to the ligand desolvation penalty, which must be greater than zero for chemically reasonable geometries, the matrix $\vec{\vec{B}}$ is positive definite and the extremum of ΔG_{var} is a minimum.[5] From Eq. (58) the optimum values of the multipoles, $\vec{Q}^{opt}$ and the minimum variational binding energy, ΔG_{var}^{opt} are obtained,

$$\vec{Q}^{opt} = -\frac{1}{2}\vec{\vec{B}}^{-1}\vec{A} \quad (59)$$

$$\Delta G_{var}^{opt} = -\frac{1}{4}\vec{A}^T\vec{\vec{B}}^{-1}\vec{A}. \quad (60)$$

ΔG_{var}^{opt} is always negative because $\vec{\vec{B}}^{-1}$ is also positive definite.

To solve for the optimal multipole distribution with the monopole (total charge) fixed ($Q_1=\mathcal{Q}$), the equation for the remaining optimal multipoles ($i\neq1$) is,

$$2\sum_{j\neq1}(\beta_{ij}-\delta_{ij}\gamma_i)Q_j^{opt}+(2\beta_{i1}\mathcal{Q}+\alpha_i)=0 \quad (61)$$

which is analogous to Eq. (59).

The above matrix equations, with the dimension truncated at $i_{max}=(l_{max}+1)^2$, can be solved numerically by relatively modest computational resources. In practice, since the α_i and β_{ij} contain a summation over an infinite number of terms, a second cutoff value of l_{cut} must be used to truncate the innermost sum in Eqs. (25) and (46). When l_{max} and l_{cut} are sufficiently large, ΔG_{var}^{opt} converges and the incremental advantage of including more multipoles essentially vanishes.

For any given receptor and geometry, we have thus described a method to determine the charge distribution of the tightest binding ligand as a set of multipoles. The deviation of the binding free energy from the optimum for any test ligand can be calculated by subtracting Eq. (60) from Eq. (58) and using Eq. (59) to eliminate $\vec{A}$,

TABLE I

$C(l', 1, l; m - m', m')$.[a]

	m' = 1	m' = 0	m' = −1
l = l' + 1	$\left[\frac{(l'+m)(l'+m+1)}{(2l'+1)(2l'+2)}\right]^{1/2}$	$\left[\frac{(l'-m+1)(l'+m+1)}{(2l'+1)(l'+1)}\right]^{1/2}$	$\left[\frac{(l'-m)(l'-m+1)}{(2l'+1)(2l'+2)}\right]^{1/2}$
l = l'	$-\left[\frac{(l'+m)(l'-m+1)}{2l'(l'+1)}\right]^{1/2}$	$\frac{m}{[l'(l'+1)]^{1/2}}$	$\left[\frac{(l'-m)(l'+m+1)}{2l'(l'+1)}\right]^{1/2}$
l = l' − 1	$\left[\frac{(l'-m)(l'-m+1)}{2l'(2l'+1)}\right]^{1/2}$	$-\left[\frac{(l'-m)(l'+m)}{l'(2l'+1)}\right]^{1/2}$	$\left[\frac{(l'+m+1)(l'+m)}{2l'(2l'+1)}\right]^{1/2}$

[a]from reference 4

APPENDIX REFERENCES

[1]J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (John Wiley and Sons, New York, 1975).

[2]L. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems* (MIT Press, Cambridge, Mass., 1988).

[3]M. E. Rose, J. Math. & Phys. 37, 215 (1958);

[4]M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley and Sons, New York, 1957).

[5]G. Strang, *Introduction to Applied Mathematics* (Wellesley-Cambridge Press, Wellesley, Mass., 1986).

Having now described a few embodiments of the present computer-implemented process, it should be apparent to those skilled in the art that the foregoing is merely illustrative and not limiting, having been presented by way of example only. Numerous modifications and other embodiments are within the scope of one of ordinary skill in the art and are contemplated as falling within the scope of the present process as defined by the appended claims and equivalent thereto.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 1

<210> SEQ ID NO 1
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Influenza A virus

<400> SEQUENCE: 1

Pro Lys Tyr Val Lys Gln Asn Thr Leu Lys Leu Ala Thr
1 5 10

What is claimed is:

1. A computer-implemented process for identifying properties of a ligand for binding to a target molecule in a solvent comprising the steps of:

receiving an indication of a selected shape of the ligand, defined in three dimensions, which complements a shape of a selected portion of the target molecule, defined in three dimensions;

determining a representation of a charge distribution which minimizes electrostatic contribution to binding free energy between the ligand and the target molecule in the solvent.

2. The computer-implemented process of claim 1, wherein the representation of the charge distribution is a set of multipoles.

3. The computer-implemented process of claim 1, further comprising the step of identifying a ligand having point charges that match the representation of the charge distribution.

4. The computer-implemented process of claim 2, further comprising the step of identifying a ligand having point charges that match the representation of the charge distribution.

5. The computer-implemented process of claim 1, further comprising the step of designing a combinatorial library containing ligands having point charges that match the representation of the charge distribution.

6. The computer-implemented process of claim 2, further comprising the step of designing a combinatorial library containing ligands having point charges that match the representation of the charge distribution.

* * * * *

Appendix B

Optimization of electrostatic binding free energy

Lee-Peng Lee
Departments of Chemistry and Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

Bruce Tidor[a)]
Department of Chemistry, Massachussetts Institute of Technology, Cambridge, Massachussets 02139-4307

(Received 9 December 1996; accepted 24 February 1997)

An analytic result is derived that defines the charge distribution of the tightest-binding ligand given a receptor charge distribution and spherical geometries. Using the framework of continuum electrostatics, the optimal distribution is expressed as a set of multipoles determined by minimizing the electrostatic free energy of binding. Results for two simple receptor systems are presented to illustrate applications of the theory. © *1997 American Institute of Physics.*
[S0021-9606(97)50221-2]

I. INTRODUCTION

One mechanism operating in many diseases is the undesirable action of a protein (here termed receptor) that can be arrested, at least in principle, through the tight binding of a molecular ligand (e.g., by sterically blocking the active site or by preventing a required conformational change).[1] To be effective as a drug, such a molecule must possess a number of important pharmacological activities, such as bioavailability and non-toxicity. One step in the discovery of drug molecules is the identification or design of tight-binding ligands. Ligand design is particularly difficult because opposing contributions to the free energy of binding must be properly tuned. For instance, increasing the magnitude of a point charge in a ligand can enhance its interaction with receptor (tending to favor binding), but it will also enhance its interaction with solvent in the unbound state (tending to disfavor binding). What magnitude charge should be chosen to balance these effects and produce the most favorable free energy of binding? The question can be generalized to all multipole terms of the ligand charge distribution. The charge distribution that optimally balances these effects will bind tightest to the receptor.

Here the problem of determining the ligand charge distribution binding tightest to a given receptor is addressed using continuum electrostatic theory. In Section II a solution is presented for the case in which both the free ligand and the bound complex are spherical regions of low dielectric surrounded by aqueous medium of high dielectric and the behavior of the system is governed by the Poisson equation. To facilitate an analytic solution the following assumptions are made: the ligand and receptor do not interact in the unbound state, the ligand charge distribution is the same in the bound and unbound state, and the ligand binds rigidly to the receptor with a unique orientation. The optimal charge distribution is obtained by expressing the ligand charge distribution as an arbitrary set of multipoles and minimizing the free energy of binding with respect to the multipoles. In Section III the theory is applied to a highly symmetric charge distribution devised for test purposes and to a second charge distribution, the terminus of an alpha-helix, present in some protein binding sites. Discussion and conclusions are presented in Section IV.

II. THEORY

The electrostatic free energy of binding is the difference between the electrostatic free energy in the bound and the unbound state, $\Delta G_{\text{binding}} = G^{\text{bound}} - G^{\text{unbound}}$ (see Fig. 1a). Because the dielectric model includes responses that affect the entropy as well as the enthalpy, the electrostatic energy is considered to be a free energy. Here we express the free energy of each state as a sum of Coulombic and reaction-field (hydration) terms involving the ligand (L), the receptor (R), and their interaction (L–R)

$$G^{\text{state}} = G^{\text{state}}_{\text{coul,L}} + G^{\text{state}}_{\text{coul,R}} + G^{\text{state}}_{\text{coul,L-R}} + G^{\text{state}}_{\text{hyd,L}} + G^{\text{state}}_{\text{hyd,R}} + G^{\text{state}}_{\text{hyd,L-R}}. \quad (1)$$

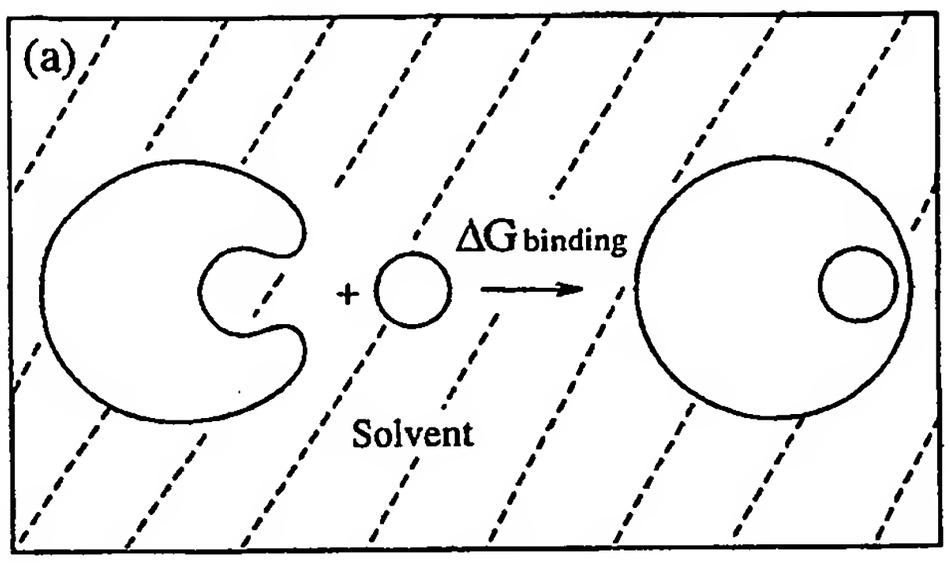
This results in the following expression for the binding free energy:

$$\Delta G_{\text{binding}} = \Delta G_{\text{coul,L-R}} + \Delta G_{\text{hyd,L-R}} + \Delta G_{\text{hyd,L}} + \Delta G_{\text{hyd,R}}, \quad (2)$$

where we have used the fact that the geometry of point charges in the receptor[2] and ligand remain fixed in the model to cancel the Coulombic self contribution of ligand and receptor and where the two L–R terms are due only to the bound state because the ligand and receptor are assumed not to interact in the unbound state. Thus, Eq. (2) describes the electrostatic binding free energy as a sum of desolvation contributions of the ligand and the receptor (which are unfavorable) and solvent-screened electrostatic interaction in the bound state (which is usually favorable). Since our goal is to vary the ligand charge distribution to optimize the electrostatic binding free energy and the last term simply adds a constant, we define a relevant variational binding energy,

$$\Delta G_{\text{var}} = \Delta G_{\text{int,L-R}} + \Delta G_{\text{hyd,L}}, \quad (3)$$

[a)]Author to whom correspondence should be addressed.



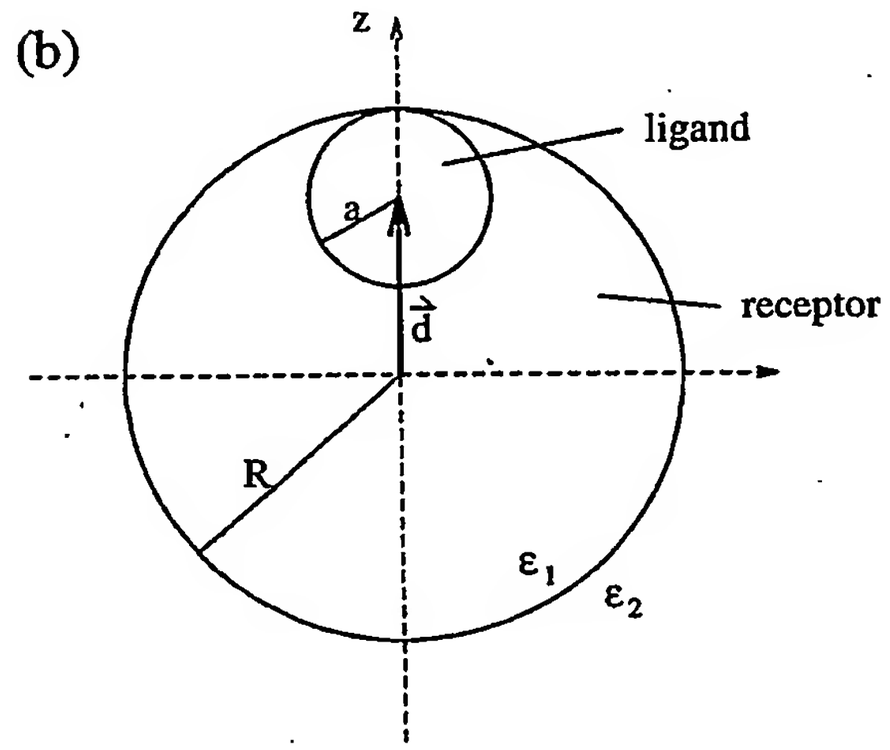


FIG. 1. Illustration of problem geometries. (a) The binding reaction is shown between a receptor (R) and spherical ligand (L) that dock rigidly to form a spherical bound-state complex. Receptor, ligand, and complex are all low-dielectric media (ϵ_1) that are surrounded by high-dielectric solvent (ϵ_2). (b) The boundary-value problem solved here involves a charge distribution in a spherical region of radius R with dielectric constant ϵ_1 surrounded by solvent with dielectric constant ϵ_2. The origin of coordinates is the center of the larger spherical region, but the charge distribution is expanded in multipoles about a point a distance d along the z-axis. The geometric requirement is that the ligand sphere not extend beyond the receptor sphere, $R \geq d+a$, although the case of equality is illustrated in the figure.

in which the first two terms on the RHS of Eq. (2) have been combined into a screened interaction term and the constant term has been dropped. Note that

$$\Delta G_{\text{int,L-R}} = \sum_{j\in R} q_j V_L^{\text{bound}}(\vec{r}_j)$$

$$= \sum_{j\in R} q_j [V_{\text{coul,L}}^{\text{bound}}(\vec{r}_j) + V_{\text{hyd,L}}^{\text{bound}}(\vec{r}_j)] \quad (4)$$

and

$$\Delta G_{\text{hyd,L}} = \frac{1}{2}\sum_{i\in L} q_i V_{\text{hyd,L}}^{\text{bound}}(\vec{r}_i) - \frac{1}{2}\sum_{i\in L} q_i V_{\text{hyd,L}}^{\text{unbound}}(\vec{r}_i), \quad (5)$$

where V_L^{state} is the total electrostatic potential in the indicated state due to the ligand charge distribution only and $V_{\text{term,L}}^{\text{state}}$ is the Coulombic or reaction-field (hydration) term, as indicated. The summations are over atomic point charges in the ligand ($i \in L$) or receptor ($j \in R$). The factor of $\frac{1}{2}$ in Eq. (5) is due to the fact that the ligand charge distribution interacts with the self-induced reaction field.

We proceed by expressing $V_{\text{coul,L}}^{\text{bound}}$, $V_{\text{hyd,L}}^{\text{bound}}$, and $V_{\text{hyd,L}}^{\text{unbound}}$, the three electrostatic potentials in Eqs. (4) and (5), in terms of the given geometry and charge distribution by solving the boundary-value problem shown in Fig. 1b. A charge distribution (corresponding to the ligand) is embedded in a sphere of radius R. We take the center of the sphere as the origin of coordinates (unprimed) but expand the charge distribution in multipoles about a second origin (primed) translated a distance d along the z-axis, so that

$$\vec{r}(r,\theta,\phi) = \vec{d}(d,\theta_d=0,\phi_d=0) + \vec{r'}(r',\theta',\phi'). \quad (6)$$

The potential everywhere satisfies the Poisson equation. Inside the sphere, it may be written as

$$V_{\text{in}}(\vec{r}) = \sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r}_i|} + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{l,m} r^l Y_{l,m}(\theta,\phi), \quad (7)$$

where the first term on the RHS is the Coulombic and the second is the reaction-field (hydration) potential, and the summation over i corresponds to the ligand point charges. Outside the sphere, the Coulombic and reaction-field potential can be combined and written as

$$V_{\text{out}}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{B_{l,m}}{r^{l+1}} Y_{l,m}(\theta,\phi), \quad (8)$$

where $A_{l,m}$ and $B_{l,m}$ are to be determined by the proper boundary conditions and $Y_{l,m}(\theta,\phi)$ are the spherical harmonics. The standard way to proceed is to expand the Coulombic term in Eq. (7) in spherical harmonics and multipoles of the charge distribution about the center of the sphere. Here we shift the origin of the multipole expansion to $\vec{d}$,

$$\sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r}_i|} = \sum_i \frac{q_i}{\epsilon_1|(\vec{r}-\vec{d})-\vec{r}_i'|} = \sum_i \frac{q_i}{\epsilon_1|\vec{r'}-\vec{r}_i'|} \quad (9)$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{2l+1} Q_{l,m}'^{*} \frac{Y_{l,m}(\theta',\phi')}{\epsilon_1 r'^{l+1}}, \quad (10)$$

where $Q'_{l,m}$ is a spherical multipole expanded about the primed origin, $\vec{d}$,

$$Q_{l,m}'^{*} \equiv \sum_i q_i r_i'^{l} Y_{l,m}^{*}(\theta_i',\phi_i'). \quad (11)$$

Note that throughout this work we adopt the definition of the $Y_{l,m}(\theta,\phi)$ used by Jackson.[3] The expression in Eq. (10) is valid for $r' > r_i'$ (i.e., outside the ligand or, more precisely, outside the sphere whose center is at $\vec{d}$ and whose radius is the distance from $\vec{d}$ to the furthest point charge).

To substitute into Eq. (7) and combine terms involving spherical harmonics, we first expand $Y_{l,m}(\theta',\phi')/r'^{l+1}$ of

Eq. (10) in terms of $Y_{l,m}(\theta,\phi)/r^{l+1}$. This is readily done using the results of Greengard,[4] which state that for $r>d$,

$$\frac{Y_{l,m}(\theta',\phi')}{r'^{l+1}} = \sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} K_{l',m',l,m}\left[\frac{4\pi(2l+1)}{(2l'+1)(2l'+2l+1)}\right]^{1/2} \times d^{l'}Y^{*}_{l',m'}(\theta_d,\phi_d)\frac{Y_{l'+l,m'+m}(\theta,\phi)}{r^{l'+l+1}}, \tag{12}$$

where

$$K_{l',m',l,m}=\left[\frac{(l'+l+m'+m)!(l'+l-m'-m)!}{(l'+m')!(l'-m')!(l+m)!(l-m)!}\right]^{1/2}. \tag{13}$$

Since we have chosen a geometry with $\theta_d=0$ (Fig. 1b), only $m'=0$ terms in Eq. (12) are non-vanishing, in which case Eq. (10) becomes

$$\sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r}_i|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{1}{\epsilon_1}\left(\frac{4\pi}{2l+1}\right)^{1/2} \times Q'^{*}_{l,m}\sum_{l'=0}^{\infty} K_{l',0,l,m}d^{l'} \times\left(\frac{4\pi}{2l'+2l+1}\right)^{1/2}\frac{Y_{l'+l,m}(\theta,\phi)}{r^{l'+l+1}}, \tag{14}$$

in which the multipole distribution is taken about the point $\vec{d}$, but the potential is expressed as a summation of spherical harmonics about the large-sphere center. The above equation can also be written as,

$$\sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r}_i|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{1}{\epsilon_1}\left(\frac{4\pi}{2l+1}\right)^{1/2}\frac{Y_{l,m}(\theta,\phi)}{r^{l+1}} \times\sum_{l'=|m|} K_{l-l',0,l',m}d^{l-l'}\left(\frac{4\pi}{2l'+1}\right)^{1/2}Q'^{*}_{l',m}, \tag{15}$$

where terms with the same $Y_{l,m}(\theta,\phi)$ are grouped together, as opposed to Eq. (14), where terms with the same $Q'^{*}_{l,m}$ are grouped.

Upon substituting Eq. (15) into Eq. (7) and matching boundary conditions at $r=R$,

$$V_{\text{in}}|_{r=R}=V_{\text{out}}|_{r=R}, \tag{16}$$

$$\epsilon_1\left.\frac{\partial V_{\text{in}}}{\partial r}\right|_{r=R}=\epsilon_2\left.\frac{\partial V_{\text{out}}}{\partial r}\right|_{r=R}, \tag{17}$$

we obtain the hydration (reaction-field) potential inside the sphere,

$$V_{\text{hyd}}(\vec{r})=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}A_{l,m}r^{l}Y_{l,m}(\theta,\phi) \tag{18}$$

$$=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\frac{4\pi}{2l+1}\right)^{1/2}r^{l}Y_{l,m}(\theta,\phi) \times\left(\frac{C_l}{R^{2l+1}}\right)\sum_{l'=|m|}^{l}K_{l-l',0,l',m}d^{l-l'} \times\left(\frac{4\pi}{2l'+1}\right)^{1/2}Q'^{*}_{l',m}, \tag{19}$$

where

$$C_l=\frac{(\epsilon_1-\epsilon_2)}{\epsilon_1[\epsilon_2+l\epsilon_1/(l+1)]}. \tag{20}$$

We can now write the various V's, with their dependence on the $Q'^{*}_{l,m}$ made explicit. $V'^{\text{bound}}_{\text{coul,L}}$ is given by Eq. (10), $V^{\text{bound}}_{\text{hyd,L}}$ is given by Eq. (19) but rewritten so that the terms with the same $Q'^{*}_{l,m}$ are collected, and $V^{\text{unbound}}_{\text{hyd,L}}$ is given by Eq. (19) with $R=a$ and $d=0$,

$$V'^{\text{bound}}_{\text{coul,L}}(\vec{r}')=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}Q'^{*}_{l,m}\frac{Y_{l,m}(\theta',\phi')}{\epsilon_1 r'^{l+1}}, \tag{21}$$

$$V^{\text{bound}}_{\text{hyd,L}}(\vec{r})=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{1/2}\left(\frac{4\pi}{2l'+1}\right)^{1/2} \times Q'^{*}_{l,m}K_{l'-l,0,l,m}d^{l'-l}\left(\frac{C_{l'}}{R^{2l'+1}}\right)r^{l'}Y_{l',m}(\theta,\phi), \tag{22}$$

$$V^{\text{unbound}}_{\text{hyd,L}}(\vec{r})=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\left(\frac{C_l}{a^{2l+1}}\right)Q'^{*}_{l,m}r^{l}Y_{l,m}(\theta,\phi). \tag{23}$$

Substituting into Eq. (4), we make explicit the dependence of $\Delta G_{\text{int,L-R}}$ on the $Q'^{*}_{l,m}$,

$$\Delta G_{\text{int,L-R}}=\sum_{j\in R}q_j[V'^{\text{bound}}_{\text{coul,L}}(\vec{r}_j)+V^{\text{bound}}_{\text{hyd,L}}(\vec{r}_j)] \tag{24}$$

$$=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}Q'^{*}_{l,m}\sum_{j\in R}q_j\left[\left(\frac{4\pi}{2l+1}\right)\frac{Y_{l,m}(\theta'_j,\phi'_j)}{\epsilon_1 r'^{l+1}_j} +\sum_{l'=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{1/2}\left(\frac{4\pi}{2l'+1}\right)^{1/2} \times K_{l'-l,0,l,m}d^{l'-l}\left(\frac{C_{l'}}{R^{2l'+1}}\right)r_j^{l'}Y_{l',m}(\theta_j,\phi_j)\right] \tag{25}$$

$$\equiv\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q'^{*}_{l,m}, \tag{26}$$

where in the last line we have defined the element $\alpha_{l,m}$, which is independent of the $Q'^{*}_{l,m}$, to be the factor multiplying $Q'^{*}_{l,m}$ in Eq. (25). Each $\alpha_{l,m}$ expresses the contribution of

a multipole to $\Delta G_{\text{int,L-R}}$ and contains all information concerning the receptor charge distribution required to obtain ΔG_{var}.

For $\Delta G_{\text{hyd,L}}$ it is useful to re-express Eq. (5) in terms of the $Q'^{*}_{l,m}$, the multipoles describing the ligand charge distribution, rather than the individual charges, q_i. We expand $V(\vec{r})$ around the center of the multipole expansion, $\vec{d}$,

$$\sum_{i\in L} q_i V(\vec{r}_i)=\sum_{i\in L} q_i V(\vec{d}+\vec{r}_i') \tag{27}$$

$$=\sum_{i\in L} q_i[V(\vec{d})+\vec{r}_i'\cdot\vec{\nabla}V(\vec{d})+\cdots]. \tag{28}$$

It has been shown by Rose that in spherical coordinates the expansion becomes,[5]

$$\sum_{i\in L} q_i V(\vec{d}+\vec{r}_i')=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{(2l+1)!!}Q'^{*}_{l,m}\mathscr{Y}_{l,m}(\vec{\nabla})V(\vec{d}), \tag{29}$$

where

$$\mathscr{Y}_{l,m}(\vec{r})\equiv r^l Y_{l,m}(\theta,\phi), \tag{30}$$

and $\mathscr{Y}_{l,m}(\vec{\nabla})$ is the operator obtained by replacing $\vec{r}$ with $\vec{\nabla}$. For positive m and when $\mathscr{Y}_{l,m}(\vec{\nabla})$ operates on a solution of the Laplace equation (i.e., $r^l Y_{l,m}(\theta,\phi)$ or $Y_{l,m}(\theta,\phi)/r^{l+1}$), it has been shown that[5]

$$\mathscr{Y}_{l,m}(\vec{\nabla})=\frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{1/2}$$
$$\times\nabla_1^m\nabla_0^{l-m}\quad\text{for}\quad m\geq 0. \tag{31}$$

The double-factorial is defined as

$$(2l+1)!!=(2l+1)\cdot(2l-1)\cdot(2l-3)\cdots 3\cdot 1 \tag{32}$$

$$=\frac{(2l+1)!}{2^l l!} \tag{33}$$

and the spherical partial derivatives are

$$\nabla_1=-\frac{1}{\sqrt{2}}(\nabla_x+i\nabla_y),\quad \nabla_{-1}=\frac{1}{\sqrt{2}}(\nabla_x-i\nabla_y),$$
$$\nabla_0=\nabla_z. \tag{34}$$

To compute $\mathscr{Y}_{l,m}(\vec{\nabla})$ for negative m, we use the fact that $Y_{l,-m}(\theta,\phi)=(-1)^m Y^{*}_{l,m}(\theta,\phi)$ and the definitions of spherical partial derivatives in Eq. (34) to obtain

$$\mathscr{Y}_{l,-m}(\vec{\nabla})=\frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{1/2}$$
$$\times\nabla_{-1}^m\nabla_0^{l-m}\quad\text{for}\quad m\geq 0. \tag{35}$$

The hydration energy of the bound ligand is then

$$G^{\text{bound}}_{\text{hyd,L}}=\frac{1}{2}\sum_{i\in L} q_i V^{\text{bound}}_{\text{hyd,L}}(\vec{d}+\vec{r}_i')$$
$$=\frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\frac{4\pi}{(2l'+1)!!}$$
$$\times Q'^{*}_{l',m'}\mathscr{Y}_{l',m'}(\vec{\nabla})V^{\text{bound}}_{\text{hyd,L}}(\vec{d}) \tag{36}$$
$$=\frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\frac{4\pi}{(2l'+1)!!}Q'^{*}_{l',m'}$$
$$\times\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l''=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{1/2}\left(\frac{4\pi}{2l''+1}\right)^{1/2}$$
$$\times Q'^{*}_{l,m}K_{l''-l,0,l,m}d^{l''-l}\left(\frac{C_{l''}}{R^{2l''+1}}\right)$$
$$\times\mathscr{Y}_{l',m'}(\vec{\nabla})(r^{l''}Y_{l'',m}(\theta,\phi))\Big|_{\vec{r}=\vec{d}}. \tag{37}$$

To evaluate $\mathscr{Y}_{l,m}(\vec{\nabla})$ in Eq. (37), we use Eq. (31) and the gradient formula[6]

$$\vec{\nabla}(\Phi(r)Y_{l,m}(\theta,\phi))$$
$$=-\left(\frac{l+1}{2l+1}\right)^{1/2}\left(\frac{d\Phi(r)}{dr}-\frac{l}{r}\Phi(r)\right)T_{l,l+1,m}(\theta,\phi)$$
$$+\left(\frac{l}{2l+1}\right)^{1/2}\left(\frac{d\Phi(r)}{dr}+\frac{l+1}{r}\Phi(r)\right)T_{l,l-1,m}(\theta,\phi), \tag{38}$$

where

$$T_{l,l',m}(\theta,\phi)\equiv\sum_{m'\in\{-1,0,1\}}\mathscr{C}(l',1,l;m-m',m')$$
$$\times Y_{l',m-m'}(\theta,\phi)\hat{\xi}_{m'}, \tag{39}$$

the $\mathscr{C}(l',1,l;m-m',m')$ are the vector addition (or Clebsch–Gordon) coefficients frequently encountered in the study of angular momentum shown in Table I,[6] and $\hat{\xi}_{m'}$ are spherical unit vectors,

$$\hat{\xi}_1=-\frac{1}{\sqrt{2}}(\hat{x}+i\hat{y}),\quad \hat{\xi}_{-1}=\frac{1}{\sqrt{2}}(\hat{x}-i\hat{y}),\quad \hat{\xi}_0=\hat{z}. \tag{40}$$

It is straightforward to show that

$$\vec{\nabla}=\hat{x}\nabla_x+\hat{y}\nabla_y+\hat{z}\nabla_z=-\hat{\xi}_1\nabla_{-1}-\hat{\xi}_{-1}\nabla_1+\hat{\xi}_0\nabla_0. \tag{41}$$

From Eqs. (38) through (41), we have

$$\nabla_\mu(r^l Y_{l,m}(\theta,\phi))$$
$$=(-1)^\mu[l(2l+1)]^{1/2}$$
$$\times\mathscr{C}(l-1,1,l;m+\mu,-\mu)r^{l-1}Y_{l-1,m+\mu}(\theta,\phi). \tag{42}$$

Using Table I, Eq. (31), and Eqs. (37) through (42), we obtain the following intermediate results:

TABLE I. $\mathscr{C}(l',1,l;m-m',m')$.[a]

	$m'=1$	$m'=0$	$m'=-1$
$l=l'+1$	$\left[\frac{(l'+m)(l'+m+1)}{(2l'+1)(2l'+2)}\right]^{1/2}$	$\left[\frac{(l'-m+1)(l'+m+1)}{(2l'+1)(l'+1)}\right]^{1/2}$	$\left[\frac{(l'-m)(l'-m+1)}{(2l'+1)(2l'+2)}\right]^{1/2}$
$l=l'$	$-\left[\frac{(l'+m)(l'-m+1)}{2l'(l'+1)}\right]^{1/2}$	$\frac{m}{[l'(l'+1)]^{1/2}}$	$\left[\frac{(l'-m)(l'+m+1)}{2l'(l'+1)}\right]^{1/2}$
$l=l'-1$	$\left[\frac{(l'-m)(l'-m+1)}{2l'(2l'+1)}\right]^{1/2}$	$-\left[\frac{(l'-m)(l'+m)}{l'(2l'+1)}\right]^{1/2}$	$\left[\frac{(l'+m+1)(l'+m)}{2l'(2l'+1)}\right]^{1/2}$

[a]From Reference 6.

$$\nabla_0^{l'-m'}(r^{l''}Y_{l'',m})=\left[\frac{(2l''+1)(l''+m)!(l''-m)!}{(2l''-2l'+2m'+1)(l''-m-l'+m')!(l''+m-l'+m')!}\right]^{1/2} r^{l''-l'+m'}Y_{l''-l'+m',m}, \tag{43}$$

$$\nabla_1^{m'}(r^{l''-l'+m'}Y_{l''-l'+m',m})=(-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''-m-l'+m')!}{2^{m'}(2l''-2l'+1)(l''-m-l'-m')!}\right]^{1/2} r^{l''-l'}Y_{l''-l',m+m'}, \tag{44}$$

$$\nabla_{-1}^{m'}(r^{l''-l'+m'}Y_{l''-l'+m',m})=(-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''+m-l'+m')!}{2^{m'}(2l''-2l'+1)(l''+m-l'-m')!}\right]^{1/2} r^{l''-l'}Y_{l''-l',m-m'}, \tag{45}$$

and the final expression for the hydration energy of the ligand in the bound state,

$$G_{\text{hyd,L}}^{\text{bound}}=\frac{1}{2}\sum_{i\in L} q_i V_{\text{hyd,L}}^{\text{bound}}(\vec{r}_i)=\frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty} Q'^{*}_{l,m}Q'_{l',m}\sum_{l''=\max(l,l')}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{1/2}\left(\frac{4\pi}{2l'+1}\right)^{1/2}$$
$$\times\frac{C_{l''}}{R^{2l''+1}}\frac{(l''+m)!(l''-m)!}{(l''-l)!(l''-l')!}\left[\frac{1}{(l+m)!(l-m)!(l'+m)!(l'-m)!}\right]^{1/2} d^{2l''-l-l'} \tag{46}$$

$$\equiv\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\beta_{l,m,l',m'}Q'^{*}_{l,m}Q'_{l',m'}, \tag{47}$$

where $\beta_{l,m,l',m'}$ is defined by the above two equations; note that $\beta_{l,m,l',m'}$ is zero for $m'\neq m$. We obtain the hydration energy of the unbound ligand by setting $d=0$ and $R=a$ in Eq. (46),

$$G_{\text{hyd, L}}^{\text{unbound}}=\frac{1}{2}\sum_{i\in L} q_i V_{\text{hyd,L}}^{\text{unbound}}(\vec{r}_i)$$
$$=\frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\left(\frac{C_l}{a^{2l+1}}\right)Q'^{*}_{l,m}Q'_{l,m} \tag{48}$$
$$\equiv\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q'^{*}_{l,m}Q'_{l,m}, \tag{49}$$

where $\gamma_{l,m}$ is defined by Eqs. (48) and (49). We write $\gamma_{l,m}$ as a function of both l and m for notational convenience, although there is no formal dependence on m.

Thus ΔG_{var} has been expressed as a function of the multipoles of the ligand charge distribution, $Q'_{l,m}$ (expanded about the center of the ligand sphere) and the elements $\alpha_{l,m}$, $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$, which do not depend on $Q'_{l,m}$. Combining Eqs. (26), (47), and (49) gives

$$\Delta G_{\text{var}}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q'^{*}_{l,m}$$
$$+\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\beta_{l,m,l',m'}Q'^{*}_{l,m}Q'_{l',m'}$$
$$-\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q'^{*}_{l,m}Q'_{l,m}. \tag{50}$$

Note that only the $\alpha_{l,m}$ depend on the receptor charges, while the $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ depend solely on the geometry of the bound and unbound states. While $\Delta G_{\text{var}}^{\text{opt}}$ is a real quantity, the $\alpha_{l,m}$ and $Q'_{l,m}$ are complex and the products $\alpha_{l,m}Q'^{*}_{l,m}$ and $Q'^{*}_{l,m}Q'_{l',m}$ involve summations over terms of the form

$Y^*_{l',m}(\theta',\phi')Y_{l,m}(\theta,\phi)$; note that the $\beta_{l,m,l'm'}$ and $\gamma_{l,m}$ are real. We rewrite $\Delta G^{\text{opt}}_{\text{var}}$ in terms of the real and imaginary parts of $\alpha_{l,m}$ and $Q'_{l,m}$,

$$\Delta G_{\text{var}}=\sum_{l=0}^{\infty}\Bigg[\alpha_{l,0}Q'_{l,0}+2\sum_{m=1}^{l}(\text{Re}\alpha_{l,m}\text{Re}Q'_{l,m}+\text{Im}\alpha_{l,m}\text{Im}Q'_{l,m})+\sum_{l=0}^{\infty}\sum_{l'=0}^{\infty}\Bigg[\beta_{l,0,l',0}Q'_{l,0}Q'_{l',0}+2\sum_{m=1}^{l}\beta_{l,m,l',m}(\text{Re}Q'_{l,m}\text{Re}Q'_{l',m}+\text{Im}Q'_{l,m}\text{Im}Q'_{l',m})\Bigg]-\sum_{l=0}^{\infty}\Bigg[\gamma_{l,0}Q'^2_{l,0}+2\sum_{m=1}^{l}\gamma_{l,m}(\text{Re}Q'^2_{l,m}+\text{Im}Q'^2_{l,m})\Bigg] \tag{51}$$

(where the summations over m are excluded for $l=0$) by noting again that $Y_{l,-m}(\theta,\phi)=(-1)^mY^*_{l,m}(\theta,\phi)$ and

$$Y^*_{l',m}(\theta',\phi')Y_{l,m}(\theta,\phi)+Y^*_{l',-m}(\theta',\phi')Y_{l,-m}(\theta,\phi)$$
$$=Y^*_{l',m}(\theta',\phi')Y_{l,m}(\theta,\phi)+Y_{l',m}(\theta',\phi')Y^*_{l,m}(\theta,\phi) \tag{52}$$
$$=2[\text{Re}Y_{l',m}(\theta',\phi')\cdot\text{Re}Y_{l,m}(\theta,\phi)+\text{Im}Y_{l',m}(\theta',\phi')\cdot\text{Im}Y_{l,m}(\theta,\phi)]. \tag{53}$$

The new variables $\text{Re}Q'_{l,m}$ and $\text{Im}Q'_{l,m}$ are re-indexed and renamed Q_i as follows:

$$\{Q'_{0,0},Q'_{1,0},\text{Re}Q'_{1,1},\text{Im}Q'_{1,1},Q'_{2,0},\text{Re}Q'_{2,1},\text{Im}Q'_{2,1},\text{Re}Q'_{2,2},\cdots\}$$
$$\leftrightarrow\{Q_1,Q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8,\cdots\}, \tag{54}$$

and similar transformations are used to create α_i, β_{ij}, and γ_i. Equation (51) can then be written as

$$\Delta G_{\text{var}}=\sum_{i=1}^{\infty}\alpha_iQ_i+\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\beta_{ij}Q_iQ_j-\sum_{i=1}^{\infty}\gamma_iQ_i^2 \tag{55}$$
$$=\sum_{i=1}^{\infty}\alpha_iQ_i+\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(\beta_{ij}-\delta_{ij}\gamma_i)Q_iQ_j, \tag{56}$$

or in matrix notation,

$$\Delta G_{\text{var}}=\vec{Q}^T\vec{\vec{B}}\vec{Q}+\vec{Q}^T\vec{A} \tag{57}$$
$$=\left(\vec{Q}+\frac{1}{2}\vec{\vec{B}}^{-1}\vec{A}\right)^T\vec{\vec{B}}\left(\vec{Q}+\frac{1}{2}\vec{\vec{B}}^{-1}\vec{A}\right)-\frac{1}{4}\vec{A}^T\vec{\vec{B}}^{-1}\vec{A}, \tag{58}$$

where $\vec{Q}$ is the vector formed by the Q_i, $\vec{A}$ is the vector formed by the α_i, $\vec{\vec{B}}$ is the symmetric matrix formed by the $(\beta_{ij}-\delta_{ij}\gamma_i)$, and completion of the square has been used to arrive at Eq. (58). Since $\vec{Q}^T\vec{\vec{B}}\vec{Q}$ in Eq. (57) corresponds to the ligand desolvation penalty, which must be greater than zero for chemically reasonable geometries, the matrix $\vec{\vec{B}}$ is positive definite and the extremum of ΔG_{var} is a minimum.[7] From Eq. (58) the optimum values of the multipoles, $\vec{Q}^{\text{opt}}$ and the minimum variational binding energy, $\Delta G^{\text{opt}}_{\text{var}}$ are obtained,

$$\vec{Q}^{\text{opt}}=-\frac{1}{2}\vec{\vec{B}}^{-1}\vec{A}, \tag{59}$$

$$\Delta G^{\text{opt}}_{\text{var}}=-\frac{1}{4}\vec{A}^T\vec{\vec{B}}^{-1}\vec{A}. \tag{60}$$

$\Delta G^{\text{opt}}_{\text{var}}$ is always negative because $\vec{\vec{B}}^{-1}$ is also positive definite.

To solve for the optimal multipole distribution with the monopole (total charge) fixed ($Q_1=\mathcal{Q}$), the equation for the remaining optimal multipoles ($i\neq1$) is

$$2\sum_{j\neq1}(\beta_{ij}-\delta_{ij}\gamma_i)Q^{\text{opt}}_j+(2\beta_{i1}\mathcal{Q}+\alpha_i)=0, \tag{61}$$

which is analogous to Eq. (59).

The above matrix equations, with the dimension truncated at $i_{\text{max}}=(l_{\text{max}}+1)^2$, can be solved numerically by relatively modest computational resources. In practice, since the α_i and β_{ij} contain a summation over an infinite number of terms, a second cutoff value of l_{cut} must be used to truncate the innermost sum in Eqs. (25) and (46). When l_{max} and l_{cut} are sufficiently large, $\Delta G^{\text{opt}}_{\text{var}}$ converges and the incremental advantage of including more multipoles essentially vanishes.

For any given receptor and geometry, we have thus described a method to determine the charge distribution of the tightest binding ligand as a set of multipoles. The deviation of the binding free energy from the optimum for any test ligand can be calculated by subtracting Eq. (60) from Eq. (58) and using Eq. (59) to eliminate $\vec{A}$,

$$\Delta G_{\text{var}}-\Delta G^{\text{opt}}_{\text{var}}=(\vec{Q}-\vec{Q}^{\text{opt}})^T\vec{\vec{B}}(\vec{Q}-\vec{Q}^{\text{opt}}). \tag{62}$$

III. RESULTS

A. Implementation

The algorithm described was implemented in a computer program whose input was l_{max} [which determined the size of the matrix in Eqs. (59) and (61)], l_{cut} [which was used to truncate the innermost summation in Eqs. (25) and (46)], the geometry of the problem, and whether the monopole of the optimum was to be free or fixed at some value. The geometry of the problem included the radius and coordinates of the center of both the bound-state and ligand spheres (on the z-axis) and the coordinates and magnitude of each partial atomic charge in the system. The dielectric constants ϵ_1 and ϵ_2 were chosen to be 4 and 80, respectively. Evaluation of the α_l, β_{ij}, and γ_i was carried out, followed by solution of the matrix equation [Eq. (59) or (61)] using LU decomposition. The eigenvalues of the $\vec{\vec{B}}$ matrix were obtained to verify that the stationary point was a minimum. All real floating-point values were represented using 64 bits. The matrix algebra was accomplished using increased-precision versions

of the appropriate subroutines given by Press *et al.*[8] The output of the program included the multipoles for the optimal charge distribution, $\Delta G_{\text{var}}^{\text{opt}}$, the nature of the stationary point, and a file recording the α_i, β_{ij} and γ_i. Typical CPU usage for a receptor with $l_{\max}=l_{\text{cut}}=40$ was 20 minutes on a Hewlett-Packard 9000/735 with the PA–7200 (99 MHz) chip, and the maximum memory used was roughly 22 MB. Because we have used a direct method (i.e., LU decomposition) to solve the matrix equation, where the matrix is of size $(l_{\max}+1)^2\times(l_{\max}+1)^2$, the time scales as $(l_{\max})^6$ and the memory scales as $(l_{\max})^4$. At this point no attempt has been made to optimize the code. For example, the matrix equation contains a particularly sparse matrix (due to the azimuthal geometry chosen for the problem) that may be used to reduce the necessary computational effort. The optimization problem may also be solved with iterative methods, such as the conjugate gradient method or various relaxation methods.

B. Test problems

The first test problem consisted of a receptor with four parallel dipolar groups, each containing a negative charge of $-0.55e$ in the $z=15.50$ plane and a positive charge of $+0.55e$ in the $z=14.25$ plane. All lengths and distances are given in units of angströms (1 Å=0.1 nm). The (x,y) coordinates of the charges were $(+1.5,+1.5)$, $(-1.5,+1.5)$, $(-1.5,-1.5)$, and $(+1.5,-1.5)$. The bound-state low-dielectric region was bounded by a sphere of radius 24.0 centered at the origin, the ligand sphere was of radius 4.0 and was centered at $(x,y,z)=(0.0,0.0,20.0)$ in the bound state.

The second test problem consisted of an idealized alpha-helix as the receptor. The helix was constructed from 18 alanine residues with acetyl and N-methylamide blocking groups at the N- and C-terminus, respectively. Coordinates were generated in the polar-hydrogen representation with the CHARMM PARAM19 (Ref. 9) bond lengths and angles and with $\phi=-57°$ and $\psi=-47°$. The partial atomic charges were adapted from the PARSE parameter set.[10] The axis of the helix coincided with the z-axis of the coordinate system and the nitrogen atom of Ala 10 was closest to the origin. The bound-state low-dielectric region was bounded by a sphere of radius 24.0 centered at the origin, the ligand sphere was of radius 4.0, and the ligand multipole distribution was centered at $(x,y,z)=(0.0,0.0,20.0)$ in the bound state (near the C-terminus of the poly-alanine alpha-helix).

C. Analysis of results

Each test problem was solved multiple times using different values of $l_{\max}$ (with l_{cut} fixed at 40 for the results shown here, though essentially indistinguishable results were obtained when the value was increased to 80) and with the monopole of the variational distribution either free or fixed at 0 or $+1e$. Figure 2 shows the convergence of the calculated $\Delta G_{\text{var}}^{\text{opt}}$ as a function of the value of $l_{\max}$ used (part a is for the four-dipolar-groups problem and part b is for the alpha-helix). In all cases the calculated $\Delta G_{\text{var}}^{\text{opt}}$ was monotonically decreasing for increasing $l_{\max}$, and for any value of $l_{\max}$, $\Delta G_{\text{var}}^{\text{opt}}$ was lower (more favorable) with free than with fixed

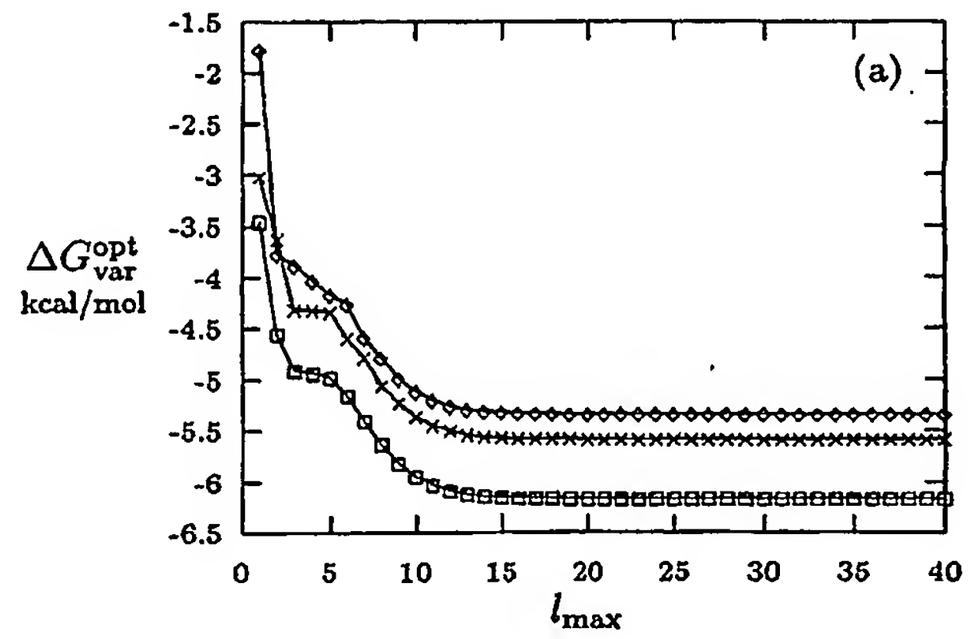
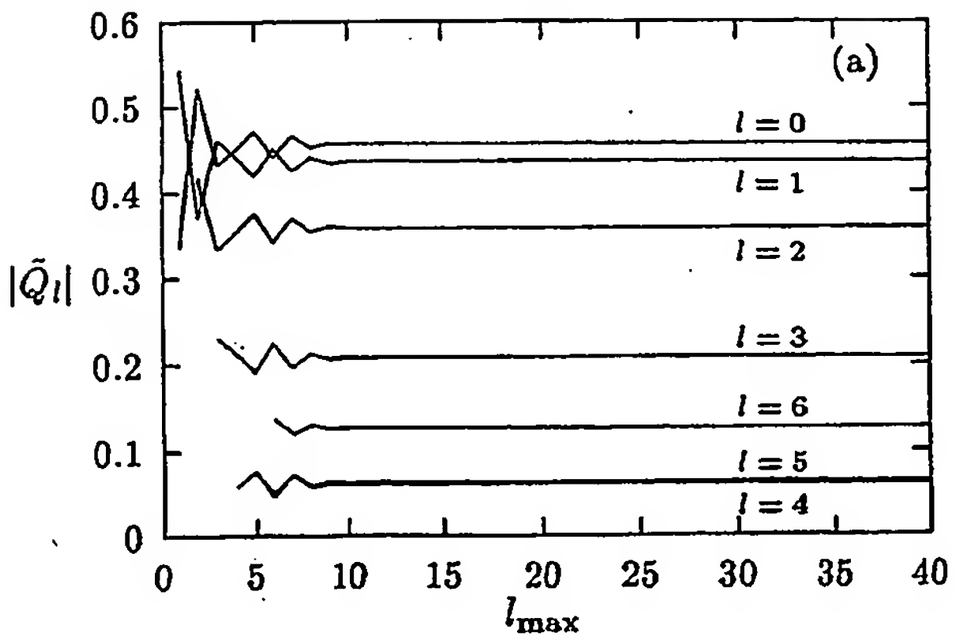


FIG. 2. Convergence of $\Delta G_{\text{var}}^{\text{opt}}$ as a function of the value of $l_{\max}$ used in the calculation. A constant value of $l_{\text{cut}}=40$ was used throughout. Optimizations in which the total charge on the ligand was free are plotted with (□), fixed at 0 with (×), and fixed at 1 with (◇) for (a) the four dipolar groups and (b) the alpha-helix.

monopole value, as expected for a variational optimization. For both test problems the value of $\Delta G_{\text{var}}^{\text{opt}}$ appeared to change very little beyond an $l_{\max}$ of 20 for floating or fixed monopole. Figure 3 shows the magnitude of the low multipole moments of the optimized distribution as a function of $l_{\max}$ (with free monopole value). The magnitude of the 2^l-pole is defined as $|\tilde{Q}_l|\equiv\sqrt{(4\pi/(2l+1))\Sigma_m(Q_{l,m}/a^l)^2}$, where a is the ligand radius. The magnitudes of the first six multipoles converged by an $l_{\max}$ of 10. Figure 4 shows the Coulombic potential due to the calculated optimal ligand, again as a function of $l_{\max}$; the potential appeared nearly converged at an $l_{\max}$ of 20. The converged Coulombic potential of the optimal ligand, plotted in the xy-plane just outside the ligand (at $z=16.0$) and computed with an $l_{\max}$ of 40, is shown in Fig. 5a for the four-dipolar-groups problem and Fig. 5c for the alpha-helix. The optimal ligand's potential contained the appropriate four-fold symmetry to match that of the four-dipolar-groups receptor, indicating that such a ligand would interact equally with all four dipoles. However, for the alpha-helix, which presents a coil of dipolar groups



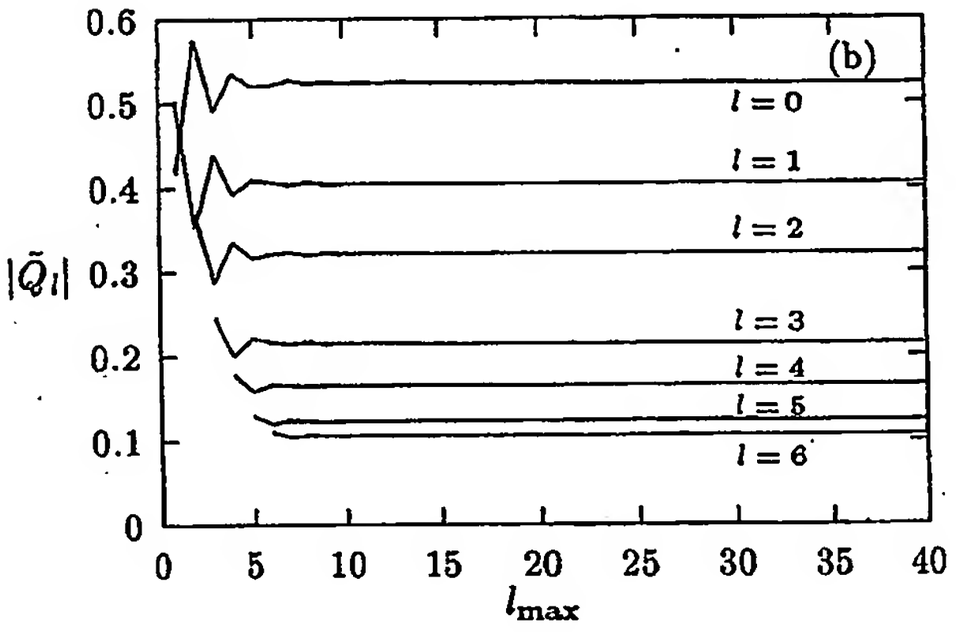
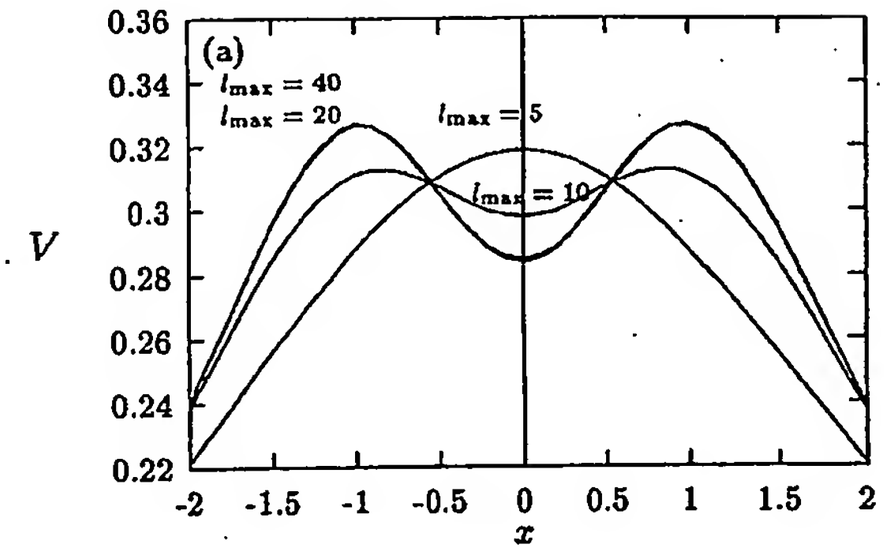


FIG. 3. Convergence of the magnitude of the lowest seven 2^l-poles for the optimal ligand as a function of the value of l_{max} used in the calculation for (a) the four dipolar groups and (b) the alpha-helix. The optimizations were performed with no constraint on the total ligand charge. In (a) note that the $l=4$ and $l=5$ lines fall nearly on top of one another.



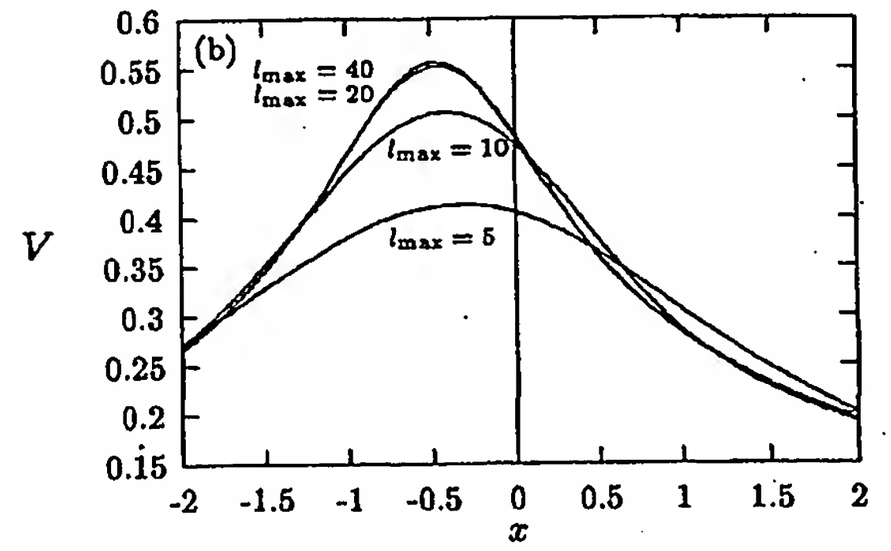


FIG. 4. Convergence of the Coulombic potential due to the optimal ligand, plotted along the line ($y=-1.1, z=16.0$) for a range of values of l_{max}, for (a) the four dipolar groups and (b) the alpha-helix. The optimizations were performed with no constraint on the total ligand charge. Note that the curves for $l_{max}=20$ and 40 are nearly identical.

receding in the z-direction, it appears that the optimal ligand computed in this manner would interact strongly only with the closest dipolar group. It is also interesting to note that the Coulombic potential due to the optimized multipole distribution calculated in this way is not a simple reflection of the Coulombic potential for the isolated receptor. Compare, for instance, the coulombic potentials due to the optimized ligand (Fig. 5a) and due to the receptor (Fig. 5b) for the four-dipolar-groups problem, both computed in the $z=16.0$ plane. The peaks in the ligand potential are "inside" those of the receptor potential. This may turn out to be a general feature of electrostatically optimized binding interactions, which are fundamentally asymmetrical, since one distribution is fixed while the other is optimized.

IV. DISCUSSION AND CONCLUSION

Analytic solutions to the Poisson equation have been used to define the multipole distribution of the ligand that produces a minimum for the free energy of binding a spherical ligand to an invariant receptor to form a spherical complex. An algorithm has been developed and implemented using numerical computation to evaluate the analytic theory, and results have been presented for two test cases. In all solutions examined to date, second-derivative analysis has verified that the stationary point is a minimum. In this sense, the multipole distribution is said to be an optimum. An important feature of the theory presented is that, by expressing the optimum as a multipole distribution, it can be solved for directly, without resorting to stochastic searches or other non-deterministic methods of optimization. This characterization of the multipole properties of the optimal charge distribution for a given spherical ligand shape and binding geometry may be useful in understanding complementary interactions in molecular binding and recognition. Such properties may prove particularly applicable to the field of ligand design either by facilitating the construction of individual tight-binding ligands or by providing descriptors that can be used to search compound libraries or aid in the design of combinatorial libraries.

The observation that an optimum can be defined within the continuum model presented here so as to provide the greatest excess of favorable interactions between ligand and receptor over unfavorable ligand desolvation energy suggests that the successful design of a tight-binding ligand may involve substantially more than the construction of a complementary-shaped molecule that provides compensating interactions for polar and charged groups in the receptor

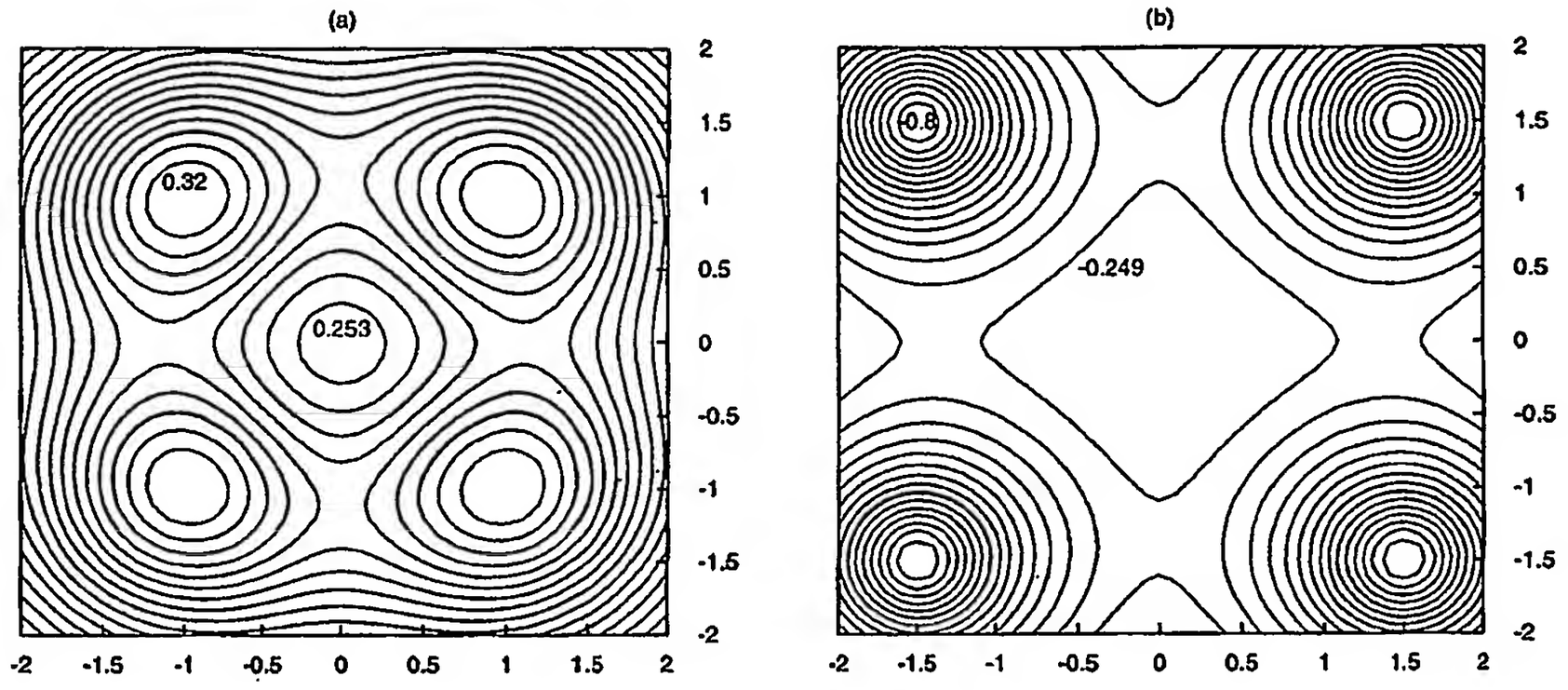


FIG. 5. Contour plot of the Coulombic potential in the $z=16.0$ plane for (a) the optimum ligand for the four dipolar groups, (b) the four dipolar groups themselves, (c) the optimum ligand for the alpha-helix, and (d) the alpha-helix itself. The optimizations were performed with no constraint on the total ligand charge. Each plot consists of equally spaced contour levels. Each label marks the closest contour level and is valid to three decimal places (i.e., 0.32 in (a) is 0.320 and −0.8 in (b) is −0.800), except −1.1 in (d), which is −1.090 but was rounded for clarity in the figure.

binding site. For example, the electrostatics of compensating a neutral, polar carbonyl group in a receptor with a neutral, polar hydroxyl may be substantially different than complementing it with a positively charged ammonium group. Moreover, due to the effects of longer-range electrostatic interactions, merely discussing the problem in terms of individually compensating pairs of groups may be inappropriate, since each group affects the overall multipole moments of the ligand. To help answer these questions, we are currently studying algorithms for designing sets of point charges, as well as molecules, that have multipole moments corresponding closely to the optimum defined by this algorithm.

The properties of the optimal multipole distribution and binding energy are worthy of further study. Here we note that ΔG^{opt}_{var} is always negative (favorable), but that the overall binding free energy, $\Delta G_{binding}$, may or may not be positive (unfavorable) due to the receptor desolvation energy.[11] Moreover, it is straightforward to prove that the magnitude of the screened ligand–receptor interaction free energy is twice that of the ligand desolvation energy at the optimum ($\Delta G^{opt}_{int,L\text{-}R} = -2\Delta G^{opt}_{hyd,L}$, so $\Delta G^{opt}_{var} = -\Delta G^{opt}_{hyd,L} = \frac{1}{2}\Delta G^{opt}_{int,L\text{-}R}$), and that the same relationship holds for the contribution of each multipole component, Q_i^{opt}.[12] Finally, the relationship between the Coulombic potential of the optimized charge distribution and that of the receptor reveals non-trivial features that reflect subtleties of how best to achieve favorable interactions in the bound state relative to ligand desolvation. For the example involving four dipolar groups, this suggests that chemical groups compensating each dipole should lie closer to the azimuthal axis than the corresponding receptor dipole.

The current theory provides a useful starting point for further studies. We are presently investigating extensions to solve the linearized[13] and the non-linear Poisson–Boltzmann equation, which would allow ionic-strength effects of the aqueous medium to be included. Moreover, it may be possible to release the restrictions that both the unbound ligand and the bound complex have spherical geometry, that the charge distribution of the ligand be the same in the bound and unbound states, and that titratable groups be treated in a fixed protonation state. It should be noted that there is no restriction in the current theory on the shape or charge distribution of the unbound receptor, since its contribution is a constant that has been eliminated in the definition of ΔG_{var}.

ACKNOWLEDGMENTS

The authors thank Moungi G. Bawendi, Christopher C. Cummins, Rick L. Danheiser, Robert W. Field, Cristina Jarque, Erik Kangas, Whay C. Lee, Stephen J. Lippard, Irwin Oppenheim, Carl O. Pabo, Robert J. Silbey, and particularly Sara E. Dempster for helpful discussions. This work was supported by the National Institutes of Health (GM47678) and the MIT Science Partnership Fund.

[1] M. Perutz, *Protein Structure: New Approaches to Disease and Therapy* (Freeman, New York, 1992).
[2] The charge distribution for the receptor need not be the same in the bound and unbound states. If they are different, this adds a constant to $\Delta G_{\text{binding}}$ that can be dropped in defining ΔG_{var} in Eq. (3).
[3] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975).
[4] L. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems* (Massachusetts Institute of Technology, Cambridge, 1988).
[5] M. E. Rose, J. Math. Phys. 37, 215 (1958).
[6] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[7] G. Strang, *Introduction to Applied Mathematics* (Wellesley–Cambridge, Wellesley, 1986).
[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1988).
[9] B. R. Brooks *et al.*, J. Comput. Chem. 4, 187 (1983).
[10] D. Sitkoff, K. Sharp, and B. Honig, J. Phys. Chem. 98, 1978 (1994).
[11] Z. S. Hendsch and B. Tidor (unpublished).
[12] S. E. Dempster, L.-P. Lee, and B. Tidor (unpublished).
[13] J. G. Kirkwood, J. Chem. Phys. 2, 351 (1934).

Claims

1. A method of modulating the antigen-binding affinity of an antibody comprising,

determining a spatial representation of an optimal charge distribution of the amino acids of the antibody and associated change in binding free energy of the antibody when bound to an antigen in a solvent;

identifying at least one candidate amino acid residue position of the antibody to be modified to alter the binding free energy of the antibody when bound to the antigen; and

selecting an elected amino acid residue for substitution for said amino acid position,

such that upon substitution, the antigen-binding affinity of the antibody is modulated.

2. The method of claim 1, further comprising substituting the elected amino acid residue at the candidate amino acid residue position.

3. A method of modulating the antigen-binding affinity of an antibody comprising,

determining a spatial representation of an optimal charge distribution of the amino acids of the antibody and associated change in binding free energy of the antibody when bound to an antigen in a solvent;

identifying at least one candidate amino acid residue position of the antibody to be modified to alter the binding free energy of the antibody when bound to the antigen;

selecting an alteration for said amino acid position,

such that upon alteration, the antigen-binding affinity of the antibody is modulated.

4. The method of claim 3, wherein the alteration is selected from the group consisting of a deletion, an insertion, and an alteration of side chain chemistry.

5. The method of claim 1 or 3, further comprising calculating the change in the free energy of binding of the antibody containing the modified amino acid or alteration when bound to the antigen, as compared to the unmodified antibody when bound to the antigen.

6. The method of claim 5, wherein the calculating step first comprises modeling the modification or alteration of the antibody *in silico*, and then calculating the change in free energy of binding.

7. The method of claim 6, wherein the calculating step uses at least one determination selected from the group consisting of a determination of the electrostatic binding energy

using a method based on the Poisson-Boltzmann equation, a determination of the van der Waals binding energy, and a determination of the binding energy using a method based on solvent accessible surface area.

8. The method of claim 1 or 3, further comprising expressing the modified or altered antibody.

9. The method of claim 1 or 3, wherein the modulation is selected from the group consisting of an increase in antibody/antigen binding affinity and a decrease in antibody/antigen binding affinity.

10. The method of claim 1, wherein the elected amino acid is from a subset of amino acids having characteristic side chain chemistry, said subset of amino acids selected from the group consisting of uncharged polar amino acid residues, nonpolar amino acid residues, positively charged amino acid residues, and negatively charged amino acid residues.

11. The method of claim 1, wherein the elected amino acid residue increases the free energy of binding between antibody and antigen when bound in a solvent, thereby decreasing antibody-antigen binding affinity.

12. The method of claim 1, wherein the elected amino acid residue decreases the free energy of binding between antibody and antigen when bound in a solvent, thereby increasing antibody-antigen binding affinity.

13. A method of modulating the antigen-binding affinity of an antibody comprising,

determining a spatial representation of an optimal charge distribution of the amino acids of the antibody and associated change in binding free energy of the antibody when bound to an antigen in a solvent,

identifying at least one candidate amino acid residue position of the antibody to be modified to alter the binding free energy of the antibody when bound to the antigen;

selecting an elected amino acid residue for substitution at said amino acid position;

modeling the elected amino acid residue for substitution *in silico*, calculating the change in free energy of binding of the modified antibody when bound to the antigen; and

substituting the elected amino acid residue for the candidate amino acid residue position such that the antigen-binding affinity of the antibody is modulated.

14. The method of claim 13, wherein the calculating step uses at least one determination selected from the group consisting of a determination of the electrostatic binding energy using a method based on the Poisson-Boltzmann equation, a determination of the van der Waals binding energy, and a determination of the binding energy using a method based on solvent accessible surface area.

15. The method of claim 13, further comprising expressing the modified antibody.

16. The method of any one of claims 1, 3, or 13, wherein in the method is repeated at least one time.

17. The method of any one of claims 1 or 3, wherein in the method is conducted *in silico*.

18. The method of any one of claims 1, 3, or 13, wherein at least one step is informed by three-dimensional structural data.

19. The method of any one of claims 1, 3, or 13, wherein at least one step is informed by data selected from the group consisting of binding data derived from an expressed antibody binding to an antigen in a solvent, crystal structure data of an antibody, crystal structure data of an antibody bound to an antigen, three-dimensional structural data of an antibody, NMR structural data of an antibody, and computer-modeled structural data of an antibody.

20. The method of any one of claims 8 or 15, wherein expressing the modified antibody is in an expression system selected from the group consisting of an acellular extract expression system, a phage display expression system, a prokaryotic cell expression system, and a eukaryotic cell expression system.

21. The method of claim 1, wherein the antibody, or antigen-binding fragment thereof, is modified at one or more positions within a CDR region(s) selected from the group consisting of V_H CDR1, V_H CDR2, V_H CDR3, V_L CDR1, V_L CDR2, and V_L CDR3.

22. The method of claim 1, wherein the antibody, or antigen-binding fragment thereof, is selected from the group consisting of an antibody, an antibody light chain (VL), an antibody heavy chain (VH), a single chain antibody (scFv), a F(ab')2 fragment, a Fab fragment, an Fd fragment, and a single domain fragment.

23. The method of claim 1, wherein the antigen-binding affinity of the antibody is predicted to be increased by a factor of about 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 5, 8, 10, 50, 10^2, 10^3, 10^4, 10^5, or 10^6, 10^7, or 10^8.

24. The method of claim 1, wherein the antigen-binding affinity of the antibody is predicted to be decreased by a factor of about 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 5, 8, 10, 50, 10^2, 10^3, 10^4, 10^5, or 10^6, 10^7, or 10^8.

25. The method of claim 1, wherein the antigen-binding affinity is determined in the presence of an aqueous solvent containing salt.

26. The method of claim 25, wherein the solvent comprises physiological concentrations of salt.

27. An antibody, or antigen-binding fragment thereof, produced by the method of any one of claims 1, 3 or 13.

28. An antibody, or antigen-binding fragment thereof, affinity matured according to the method of any one of claims 1, 3 or 13.

29. A plurality of antibodies, or antigen-binding fragments thereof, produced by the method of any one of claims 1, 3 or 13.

30. A nucleic acid encoding the antibody, or antigen-binding fragment thereof, of claim 27.

31. A host cell encoding the nucleic acid of claim 30.

32. An antibody, or binding fragment thereof, produced by culturing the host cell of claim 31 under conditions such that antibody, or binding fragment thereof, is expressed.

33. A pharmaceutical composition comprising the antibody, or antigen-binding fragment thereof, of claim 27.

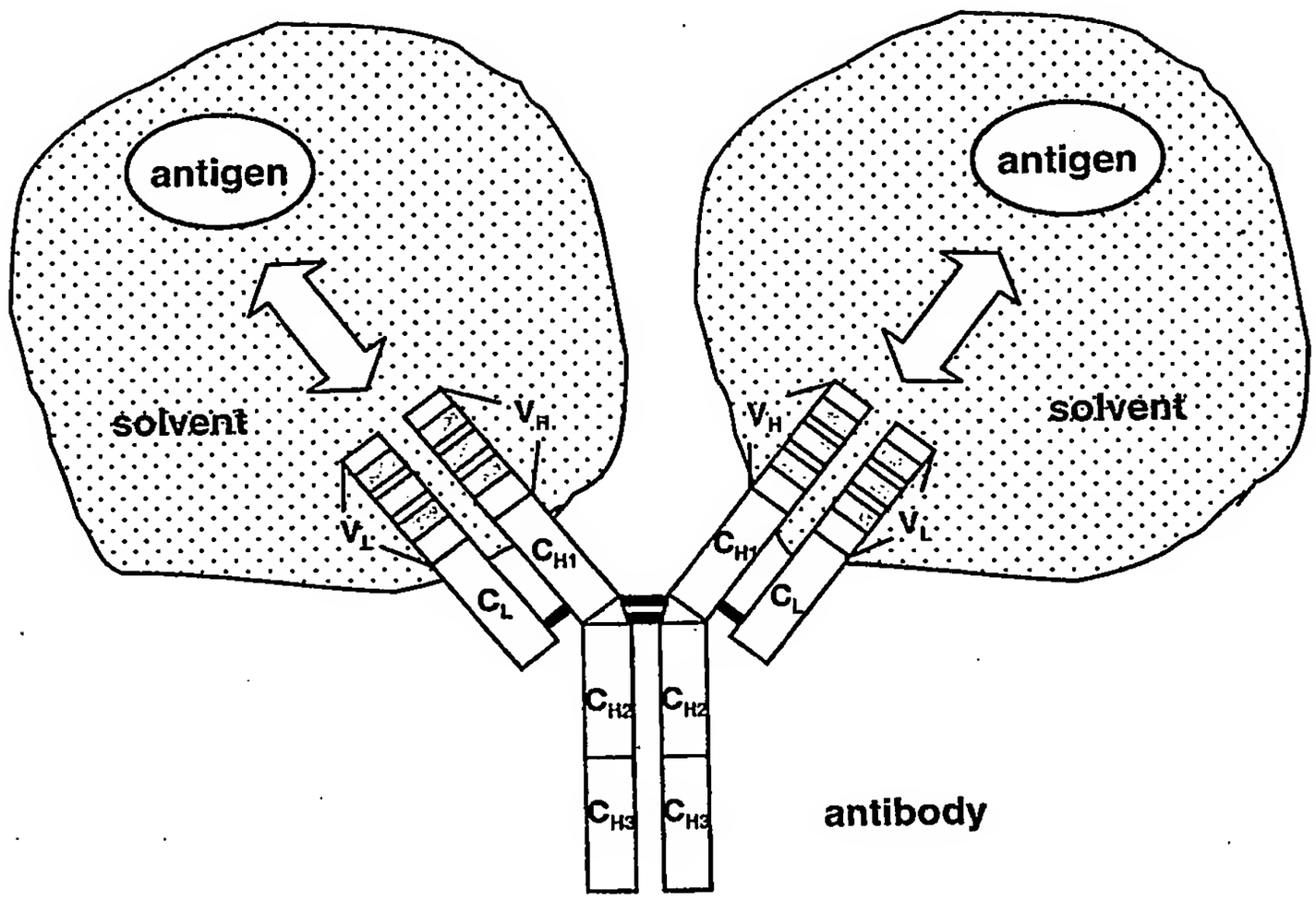
34. A method for treating or preventing a human disorder or disease comprising, administering a therapeutically-effective amount of the pharmaceutical composition of claim 33, such that therapy or prevention of the human disease or disorder is achieved.

35. The method of any one of claims 1, 3 or 13, wherein one or more steps is computer-assisted.

36. A medium suitable for use in an electronic device having instructions for carrying out one or more steps of the method of any one of claims 1, 3 or 13.

37. A device for carrying out one or more steps of the method of any one of claims 1, 3 or 13.

Fig. 1


antigen
antigen
solvent
solvent
V_H
V_H
V_L
V_L
C_{H1}
C_{H1}
C_L
C_L
C_{H2}
C_{H2}
C_{H3}
C_{H3}
antibody

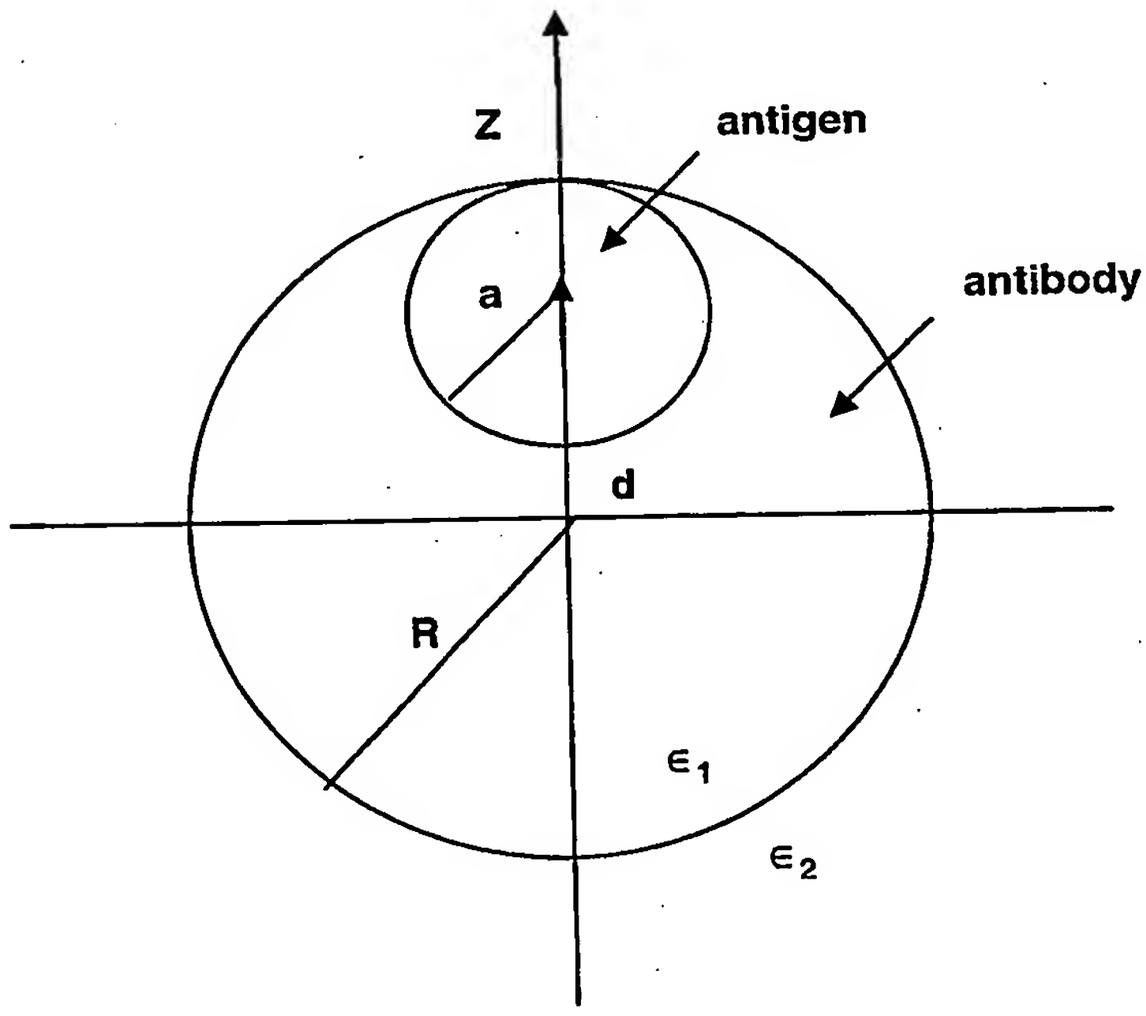

Z
antigen
antibody
a
d
R
ϵ_1
ϵ_2

Fig. 2

5C8 light variable

```
          D  I  V  L  T  Q  S  P  A  T  L  S  V  S  P  G  E  R  A  T  I  S  C  R  A  S  Q·
 952      GACATCGTTC TCACACAGTC TCCTGCTACC TTATCTGTAT CTCCGGGAGA GAGGGCCACC ATCTCATGCA GGGCCAGCCA
          ·R  V  S  S  S  T  Y  S  Y  M  H  W  Y  Q  Q  K  P  G  Q  P  P  K  L  L  I  K  Y·
1032      ACGTGTCAGT TCATCTACCT ATAGTTATAT GCACTGGTAC CAACAGAAAC CAGGACAGCC ACCCAAACTC CTCATCAAGT
          ·  A  S  N  L  E  S  G  V  P  A  R  F  S  G  S  G  S  G  T  D  F  T  L  T  I  S
1112      ATGCATCCAA CCTAGAATCT GGGGTCCCTG CCAGGTTCAG TGGCAGTGGG TCTGGGACTG ACTTCACCCT CACCATCTCT
          S  V  E  P  E  D  F  A  T  Y  Y  C  Q  H  S  W  E  I  P  P  T  F  G  G  G  T  K·
1192      TCTGTGGAGC CGGAGGATTT TGCAACATAT TACTGTCAGC ACAGTTGGGA GATTCCTCCG ACGTTCGGTG GAGGGACCAA
          ·L  E  I  K  (SEQ ID NO: 2)
1272      GCTCGAGATT AAG (SEQ ID NO: 1)
```

5C8 heavy variable

```
          Q  V  Q  L  V  Q  S  G  A  E  V  V  K  P  G  A  S  V  K  L  S  C  K  A  S  G  Y·
 189      CAGGTTCAGC TGGTGCAGTC AGGGGCTGAA GTGGTGAAGC CTGGGGCTTC AGTGAAGTTG TCCTGCAAGG CTTCTGGCTA
          ·I  F  T  S  Y  Y  M  Y  W  V  K  Q  A  P  G  Q  G  L  E  W  I  G  E  I  N  P  S·
 269      CATCTTCACC AGTTATTATA TGTACTGGGT GAAGCAGGCG CCCGGACAAG GCCTTGAGTG GATTGGAGAG ATTAATCCTA
          ·N  G  D  T  N  F  N  E  K  F  K  S  K  A  T  L  T  V  D  K  S  A  S  T  A  Y
 349      GCAATGGTGA TACTAACTTC AATGAGAAGT TCAAGAGTAA GGCCACACTG ACTGTAGACA AATCCGCCAG CACAGCATAC
          M  E  L  S  S  L  R  S  E  D  T  A  V  Y  Y  C  T  R  S  D  G  R  N  D  M  D  S·
 429      ATGGAGCTCA GCAGCCTGAG GTCTGAGGAC ACTGCGGTCT ATTACTGTAC AAGATCGGAC GGTAGAAATG ATATGGACTC
          ·W  G  Q  G  T  L  V  T  V  S  S  (SEQ ID NO: 4)
 509      CTGGGGCCAA GGGACCCTGG TCACCGTCTC CTCA (SEQ ID NO: 3)
```